

البيانات الكبرى

ثورة

ستغير الطريقة التي
نعيش، ونعمل، ونفكر بها

فيكتور ماير- شونبرجر
كينيث كوكير

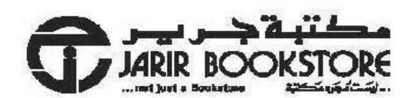


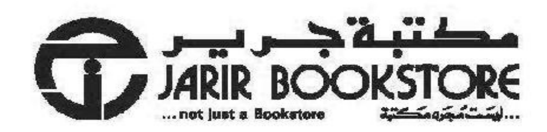
البيانات الكبرى

ثورة

ستغير الطريقة التي نعيش، ونعمل، ونفكر بها

فيكتور ماير - شونبرجر
كينيث كوكير





للتعرف على فروعنا في
المملكة العربية السعودية - قطر - الكويت - الإمارات العربية المتحدة
نرجو زيارة موقعنا على الإنترنت www.jarirbookstore.com
للمزيد من المعلومات الرجاء مراسلتنا على: jbpublications@jarirbookstore.com

الطبعة الأولى ٢٠١٤

المملكة العربية السعودية ص.ب. ٣١٩٦ الرياض ١١٤٧١ - تليفون: ٩٦٦١١٤٦٢٦٠٠٠ + - فاكس: ٩٦٦١١٤٦٥٦٣٦٣ +

A Revolution That

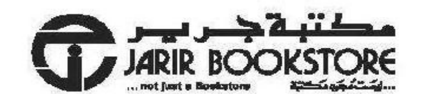
BIG

Will Transform How We

DATA

Live, Work, and Think

VIKTOR MAYER-SCHÖNBERGER
and KENNETH CUKIER



1
الآن

في عام 2009، تم اكتشاف فيروس جديد من فصيلة فيروس الإنفلونزا، يشمل خصائص كل من فيروسي إنفلونزا الطيور وإنفلونزا الخنازير مجتمعين. انتشرت هذه الفصيلة الجديدة، التي أُطلق عليها H1N1، سريعًا. وفي غضون أسابيع، خشيت الوكالات الصحية العامة في جميع أنحاء العالم من حدوث تفشٍّ وبائي خطير. حذر بعض المعلقين من تفش وبائي يشبه ما حدث عام 1918 من تفشي مرض الإنفلونزا الإسبانية الذي أصاب نصف مليار شخص وقتل عشرات الملايين. الأمر الأسوأ، أنه لا يوجد ترياق جاهز لهذا الفيروس الجديد، وكان الأمل الوحيد أمام السلطات الصحية هو أن تبطئ من انتشاره، ولكن لكي تفعل ذلك كانت تحتاج في البداية إلى معرفة مكانه.

في الولايات المتحدة الأمريكية، طلبت مراكز مكافحة الأمراض والوقاية منها من الأطباء أن يبلغوهم بحالات الإنفلونزا الجديدة، ولكن المعلومات عن الوباء عادة ما كانت تأتي متأخرة لأسبوع أو اثنين. قد يشعر الناس بالمرض لأيام قبل أن يفكروا في استشارة الطبيب. كان نقل المعلومات للمؤسسات الصحية يستغرق وقتًا, وكانت مراكز مكافحة الأمراض والوقاية منها تحدد أعداد المرضى مرة واحدة في الأسبوع. مع مرض سريع الانتشار مثل هذا المرض، تعتبر فترة الأسبوعين كأنها الدهر بأكمله، وتسبب هذا التأخير في إصابة الوكالات الصحية العامة بالعمى في أكثر اللحظات حرجًا.

في الحقيقة، وقبل تصدر فيروس H1N1 عناوين الصحف بأسابيع قليلة، نشر المهندسون في شركة الإنترنت العملاقة جوجل ورقة بحثية مميزة في الجريدة العلمية المسماة *Nature* أثارت زوبعة بين جموع المسئولين عن الصحة العامة وعلماء الحاسب ولكنهم تجاهلوها رغمًا عن ذلك. وضح المؤلفان كيف "توقعت" شركة جوجل انتشار إنفلونزا الشتاء في الولايات المتحدة الأمريكية، ليس فقط على مستوى الدولة بأكملها، بل أيضًا مع تحديد مناطق بعينها وربما ولايات بأكملها. تمكنت الشركة من فعل هذا الأمر من خلال متابعة ما يبحث عنه الناس عبر الإنترنت. ولأن شركة جوجل يصلها يوميًّا ما يزيد على 3 مليارات من طلبات البحث وتحتفظ بها جميعًا، كانت تمتلك كمًّا هائلاً من البيانات لتعمل من خلاله.

أخذت جوجل أكثر 50 مليون بحث شيوعًا قام به الأمريكيون وقارنتها ببيانات مراكز مكافحة الأمراض والوقاية منها عن انتشار الإنفلونزا الموسمية في الفترة ما بين عامي 2003 و 2008. كانت الفكرة أن يتم تحديد المناطق المصابة بفيروس الإنفلونزا من خلال ما يبحث عنه الناس على شبكة الإنترنت. حاول آخرون أن يفعلوا المثل باستخدام مصطلحات البحث عبر الإنترنت التي يكتبها الناس في محركات البحث، ولكن لم يمتلك أي منهم الكم الكافي من البيانات أو قدرة

المعالجة أو التحليل الإحصائي التي تمتلكها جوجل.

رغم أن العاملين في جوجل كانوا يخمنون بأن ما يبحث عنه الناس على محرك البحث الخاص بهم قد يهدف إلى جمع معلومات عن الإنفلونزا ــ عبارات على غرار "دواء لعلاج السعال والحمى" ــ فإن هذا لم يكن المقصود: إنهم لم يكونوا يعلمون، وصمموا نظامًا لا يولي أي اهتمام بالأمر. كل ما فعله نظامهم هو البحث عن العلاقة بين معدل تكرار بحث معين وانتشار الإنفلونزا طبقًا للزمان والمكان. في مجمل الأمر، قاموا بمعالجة 450 مليون من النماذج الرياضية المربكة المختلفة حتى يتمكنوا من اختبار مصطلحات البحث، مع مقارنة توقعاتهم بحالات الإصابة الفعلية بالإنفلونزا طبقًا لبيانات مراكز السيطرة والوقاية من الأمراض في عامي 2007 و 20. ولقد أصابوا: حيث وجد نظامهم مجموعة مكونة من 45 مصطلحًا بحثيًّا عندما يتم استخدامها مجتمعة في نموذج رياضي، تربط بشكل وثيق بين توقعاتهم والبيانات الرسمية على مستوى الدولة بأكملها. وكما كانت تفعل مراكز مكافحة الأمراض والوقاية منها، تمكنوا من تحديد أين ينتشر مرض الإنفلونزا، ولكن على العكس من تلك المراكز تمكنوا من توقع هذا على الفور ليس في خلال أسبوع أو اثنين.

خلال كارثة H1N1 في عام 2009، أثبت نظام جوجل أنه مؤشر أكثر فائدة وسرعة من الإحصائيات الحكومية التي كانت بطبيعتها بطيئة في إصدار التقارير؛ مما أمد المسئولين عن الصحة العامة بمعلومات قيمة عن الوضع.

المثير للدهشة، أن أسلوب جوجل لم يتضمن الحصول على مسحات من باطن الفم أو استشارة الأطباء، بل اعتمدت بشكل كامل على "البيانات الضخمة" ــ قدرة المجتمع على تطويع المعلومات بطرق مستحدثة لإنتاج أفكار مفيدة أو منتجات وخدمات قيمة.

وبهذا، في المرة القادمة التي سيواجه فيها العالم انتشارًا وبائيًّا أيًّا كان، سيكون مسلحًا بأداة أفضل لتوقع انتشاره ومكافحته.

إن الصحة العامة مجال واحد من بين المجالات الكثيرة التي يمكن للبيانات الضخمة أن تحدث فارقًا فيها؛ حيث إن هناك قطاعات أعمال كاملة أعادت تشكيل نفسها بفضل البيانات الضخمة. إن شراء تذاكر الطيران مثال جيد على هذا الأمر.

في عام 2003، أراد "أورين إتزيوني" السفر من سياتل إلى لوس أنجلوس لحضور حفل زفاف أخيه الأصغر. قبل عدة أشهر من حفل الزفاف، دخل "أورين" إلى موقع إحدى شركات الطيران وحجز تذكرة عبر شبكة الإنترنت، معتقدًا أنه إن حجز التذكرة مبكرًا فستكون أرخص سعرًا. عندما استقل الطائرة، دفعه فضوله لسؤال المسافر الذي يجلس في المقعد المجاور عن سعر تذكرته ومتى اشتراها. تبين أن الرجل كان قد دفع مبلغًا يقل بكثير عما دفعه "إتزيوني"، رغم أنه اشتراها قبل وقت قريب. شعر "إتزيوني" بالضيق وسأل مسافرًا آخرًا وثالثًا وهكذا، وكان أغلبهم قد دفع مبالغ تقل عما دفعه.

بالنسبة لأغلبنا، يختفي شعورنا بالخيانة المادية بمجرد أن نغلق الطاولة القابلة للطي أمامنا ونضع مقاعدنا في وضع قائم وثابت استعدادًا للإقلاع. ولكن "إتزيوني"

أحد ألمع علماء الحاسب في أمريكا، حيث يرى العالم على أنه مجموعة من المشكلات المتعلقة بالبيانات الضخمة ـــ عالج المشكلات التي بمقدوره علاجها. وهي المشكلات التي تخصص في علاجها منذ تخرج في جامعة هارفارد عام 1986 خلال تحضيره لدرجته العلمية الأولى في التخصص في مجال علوم الحاسب.

من خلال موقعه في جامعة واشنطن، قام "إتزيوني" بتأسيس عدد من شركات البيانات الضخمة قبل حتى أن يشتهر مصطلح "البيانات الضخمة"؛ حيث ساعد على إنشاء أحد محركات البحث الأولى عبر الإنترنت، ميتا كراولر، الذي أُطلق عام 1994 واشتراه موقع إنفو سبيس ليصبح فيما بعد ملكية كبيرة عبر الإنترنت. شارك أيضًا في تأسيس موقع نتبوت، والذي يعد أول موقع كبير للمقارنة التسويقية، وباعه إلى موقع إكسيت. وهناك الموقع الذي بدأه لاستخراج المعاني من النصوص، والذي كان يُطلق عليه كلير فورست، اشترته وكالة رويترز فيما بعد.

عودة إلى موضوعنا، كان "إتزيوني" مصرًّا على ابتكار طريقة تجعل الناس يدركون ما إذا كانت أسعار التذاكر التي يرونها عبر الإنترنت تعتبر صفقة جيدة أم لا. إن مقعد الطائرة ما هو إلا سلعة: كل منها لا يختلف عن غيره من المقاعد على الطائرة نفسها، ولكن تختلف أسعارها بشكل كبير طبقًا لعشرات الآلاف من العوامل التي تعلمها شركات الطيران وحدها.

استنتج "إتزيوني" أنه لا يحتاج إلى فك شفرات النظم أو المسببات التي تعمل على اختلاف أسعار التذاكر، بل كل ما كان عليه فعله هو توقع ما إذا كانت الأسعار سترتفع أو تنخفض في المستقبل. إنه أمر ممكن فعله وإن كان لا يخلو من صعوبة. كل ما كان يحتاج إليه هو تحليل جميع مبيعات التذاكر لخط جوي بعينه وفحص الأسعار المدفوعة في عدد أيام محدد قبل يوم الإقلاع.

إذا كان متوسط سعر التذكرة سينخفض، فمن المنطقي أن ينتظر المشتري ويشتري التذكرة لاحقًا، أما إذا كان متوسط سعر التذكرة يرتفع دائمًا، يقوم النظام بإسداء نصيحة للمشتري بأن يشتري التذكرة على الفور بالسعر المعروض. بمعنى آخر، كان كل ما يريده هو نسخة معدلة من الاستقصاء غير الرسمي الذي أجراه "إتزيوني" على ارتفاع 30000 قدم. دون شك، كان هذا الأمر يعد من أصعب الأمور في مجال علوم الحاسب، ولكنه قادر على فعله؛ لذا فقد بدأ العمل.

باستخدام عينة مكونة من 12 ألف حالة رصد لأسعار التذاكر التي حصل عليها "بصعوبة" من أحد مواقع الرحلات على مدار فترة وصلت إلى 41 يومًا، تمكن "إتزيوني" من إنشاء نموذج توقعي وفر على مستخدميه أموالًا طائلة. النموذج لا يقدم **تعليلًا**، بل يعنى فقط **بالنتائج**. لهذا السبب، لم يتمكن النظام من معرفة أي من المتغيرات التي تؤثر على قرارات شركات الطيران لتحديد أسعار التذاكر، مثل عدد المقاعد غير المباعة، أو المواسم أو مبيت ليلة سبت في المطار بسبب تأخر الطائرة قد يقلل من أسعار التذاكر. بنى النظام توقعاته على ما يعلمه بالفعل: الاحتمالات التي جمعها من البيانات المتوافرة عن رحلات جوية أخرى. قال "إتزيوني": "أن تشتري أو لا تشتري، هذا هو السؤال"؛ لذا أطلق على بحثه مشروع "هاملت".

تطور المشروع الصغير ليصبح شركة ذات رأس مال كبير تُدعى "فايركاست". من خلال توقع ما إذا كانت أسعار رحلات الطيران من المحتمل أن ترتفع أو تنخفض، ومدى الارتفاع أو الانخفاض الذي سيحدث، مكنت "فايركاست" المستهلكين من تقرير متى يضغطون زر "الشراء"؛ حيث سلحتهم بمعلومات لم يكونوا ليحصلوا عليها من قبل. وتقصيًا للشفافية عن عملها، قامت "فايركاست" بتقدير درجة مصداقية توقعاتها وعرضتها على مستخدميها أيضًا.

كان النظام يحتاج لكم هائل من البيانات حتى يظل يعمل. ولتحسين أداء النظام، حصل "إتزيوني" على إحدى قواعد بيانات حجز تذاكر الطيران لإحدى الشركات العاملة في المجال. بفضل هذه المعلومات، أصبح النظام قادرًا على التوقع بناءً على كل مقعد في كل رحلة طيران لأغلب الخطوط الجوية العاملة في الولايات المتحدة الأمريكية لفترة مستقبلية تصل لعام كامل. تستخدم "فايركاست" الآن حوالي 200 مليار من سجلات أسعار تذاكر الطيران لتقوم بتوقعاتها، وتعمل على توفير مبالغ طائلة على مستخدميها.

لم يكن "إتزيوني"، بشعره البني الرملي وضحكته البارزة الأسنان وملامحه الوسيمة البريئة، يبدو كمن يستطيع أن يحرم شركات الطيران من أرباح متوقعة تقدر بملايين الدولارات. في حقيقة الأمر، كان يعتزم فعل أكثر من هذا؛ ففي عام 2008 كان يخطط ليطبق النظام ذاته على بضائع أخرى، مثل غرف الفنادق وتذاكر الحفلات والسيارات المستعملة: أي بضائع ذات فرق بسيط في نوعيتها مع فارق كبير في أسعارها مع توافر أطنان من البيانات عنها. ولكن قبل أن يتمكن من تنفيذ مخططه، جاءت شركة مايكروسوفت تطرق بابه، مشترية "فايركاست" بحوالي 110 ملايين دولار، ودمجت النظام في محرك بحث "بينج". بحلول عام 2012، كان النظام يعطي توقعات صحيحة بنسبة 75% ووفر للمسافرين حوالي 50 دولارًا عن كل تذكرة.

إن "فايركاست" مثال صغير على شركات البيانات الضخمة ومثال على التوجه الذي سيسلكه العالم في المستقبل. لم يكن "إتزيوني" ليتمكن من إنشاء تلك الشركة قبل خمس أو عشر سنوات؛ حيث قال: "لقد كان هذا الأمر مستحيلاً"؛ حيث إن كمية القوة الحاسوبية والمساحة التخزينية كانت باهظة الثمن للغاية في ذلك الوقت. رغم أن التغيرات التكنولوجية من العوامل الرئيسية التي جعلت تأسيس هذه الشركة ممكنًا، فإن هناك أمرًا آخر أكثر أهمية قد تغير، أمرًا يدل على البراعة، وهو أن طريقة التفكير في الكيفية التي يمكن استغلال البيانات بها قد تغيرت هي الأخرى.

لم يعد الناس يعتبرون البيانات عبارة عن شيء ساكن أو منتهي الصلاحية، شيء تنتهي فائدته بمجرد أن يتم تحقيق الغرض الذي تم جمعها من أجله، مثل أن تحط الطائرة على الأرض (أو كما في حالة جوجل، بمجرد أن تتم معالجة طلب البحث)، بل أصبحت البيانات أصولاً ثابتة تقوم عليها شركات، ومدخلًا اقتصاديًّا حيويًّا، يُستخدم لإنشاء شكل جديد من القيمة الاقتصادية. في حقيقة الأمر، ومع وجود العقلية المناسبة، يمكن أن يُعاد استخدام البيانات بمهارة لتصبح منبعًا للابتكار

والخدمات الجديدة؛ حيث يمكنها أن تكشف الكثير لمن يملكون التواضع والاستعداد والأدوات اللازمة للاستماع لها.

دع البيانات تتحدث

يمكنك أن ترى ثمار مجتمع المعلومات بسهولة؛ حيث ستجد هاتفًا محمولاً في جيب كل شخص وجهاز حاسب في حقيبة ظهر كل شخص وكمًّا كبيرًا من أنظمة تكنولوجيا المعلومات في المكاتب الخلفية من كل حدب وصوب. ولكن الأمر الذي لا يمكن ملاحظته هو المعلومات ذاتها. بعد نصف قرن من تحول الحواسب لتصبح الاتجاه السائد في المجتمع، بدأت البيانات في التكدس لتنذر بأن شيئًا جديدًا ومميزًا في طريقه للحدوث. ليس فقط أن العالم أصبح مغمورًا بكم من المعلومات يزيد بكثير عن ذي قبل، بل إن هذه المعلومات تتزايد بشكل أسرع. إن تغير المعايير قد غير من حالة المعلومات، وإن التغيير الكمي قد تسبب في تغيير نوعي. إن العلوم على غرار الفلك والوراثة، والتي انتشرت بشدة في مطلع القرن الحادي والعشرين، هي ما شكلت مصطلح "البيانات الضخمة"، وقد اتسع المفهوم في الوقت الحالي ليشمل جميع مجالات الحياة البشرية.

لا يوجد تعريف محدد للبيانات الضخمة؛ حيث إن الفكرة في بدايتها كانت أن كم المعلومات زاد بصورة عجزت أمامها مساحات التخزين في أجهزة الحاسب عن استيعاب كمية المعلومات المراد معالجتها؛ لذا احتاج مهندسو الحاسب إلى إعادة تنظيم الأدوات التي يستخدمونها لتحليل البيانات. وكان هذا هو منشأ تكنولوجيات المعالجة الجديدة على غرار MapReduce من شركة جوجل وما يعادلها من التكنولوجيات مفتوحة المصدر مثل Hadoop من شركة ياهو. أتاحت تلك التكنولوجيات الفرصة ليتم تحليل كم من البيانات أكبر من ذي قبل، والأهم أن البيانات لم تعد تُدرج في صفوف منمقة أو جداول بيانات تقليدية. لاحت أيضًا في الأفق تكنولوجيات أخرى مستهلكة للبيانات تستغني عن الأشكال الهرمية والمتشابهة القديمة التي لم تتغير. في الوقت ذاته، ولأن شركات الإنترنت كانت قادرة على جمع كم هائل من البيانات وكانت تمتلك حافزًا ماليًّا ملحًّا لاستخراج المعنى من تلك البيانات، أصبحت من المستخدمين الأوائل لتكنولوجيات المعالجة الحديثة؛ حيث سبقت الشركات التي تعمل خارج الإنترنت والتي كانت قد سبقت شركات الإنترنت، في بعض الحالات، بعقود من الخبرة.

هناك أسلوب وحيد للتفكير في تلك المشكلة في الوقت الحالي ـــ وهو الأسلوب الذي نتبعه في هذا الكتاب: البيانات الضخمة هي الأمور التي يمكن للمرء أن يقوم بها على مقياس واسع ولا يمكنه القيام بها على مقياس أصغر؛ لاستخراج أفكار جديدة أو إنشاء أشكال جديدة من القيمة المادية، بطرق تغير من شكل الأسواق والمؤسسات والعلاقات بين المواطنين والحكومات وغيرها من الأمور.

ولكن هذه هي البداية. إن حقبة البيانات الضخمة تتحدى أسلوب حياتنا وتفاعلنا مع العالم من حولنا. المثير للدهشة، أن المجتمع سيحتاج إلى التخلص من بعض هوسه بالسببية في مقابل بعض العلاقات التبادلية البسيطة: دون أن يعلم **لماذا**، بل يعلم

فقط **ماذا**. هذا الأمر من شأنه أن يقلب رأسًا على عقب قرونًا من الممارسات الثابتة ويتحدى استيعابنا الأساسي لكيفية صنع القرار وإدراك الواقع.

إن البيانات الضخمة في غمرة تحول هائل سيحدث في المستقبل. ومثلها مثل الكثير من التكنولوجيات الحديثة، ستصبح البيانات الضخمة ضحية دورة المبالغة الشهيرة للشركات العاملة في وادي السليكون: بعد أن ظهرت على أغلفة المجلات ونوقشت في مؤتمرات المجالات المختلفة، سوف يُنبذ هذا التوجه وستتعثر الكثير من الشركات المبهورة بالعمل في مجال البيانات. ولكنْ كل من المعجبين والرافضين لا يدركون جيدًا أهمية الأمر الذي سيحدث. مثلما مكننا المنظار من فهم الكون والمجهر من فهم الكائنات الدقيقة، ستمكننا التقنيات الحديثة لجمع وتحليل كميات هائلة من البيانات من فهم عالمنا بطرق ما زلنا نتعلم كيف نقدرها. إننا في هذا الكتاب لا نبشر بقدوم البيانات الضخمة، بل نحن من أرسلتنا لنبلغ العالم بقدومها. ومرة أخرى أقول، لن تكون الثورة الحقيقية في الأجهزة التي تعالج البيانات بل في البيانات نفسها وكيفية استخدامنا إياها.

حتى نتمكن من استيعاب أن ثورة المعلومات أصبحت على وشك الحدوث، تأمل التوجهات السائدة في المجتمع الآن. إن فضاءنا الرقمي يتسع باضطراد. خذ الفلك على سبيل المثال، عندما بدأ برنامج سلوان للمسح الفضائي الرقمي عام 2000، جمع منظاره في مدينة نيوميكسيكو في بضعة أسابيع بيانات تزيد عما جُمع في تاريخ الفلك بأكمله. وبحلول عام 2010 كان أرشيف برنامج المسح الفضائي مكدسًا بحوالي 140 تيرابايت من المعلومات. ولكن المنظار الذي سيخلفه، منظار المسح الضوئي الفضائي الكبير الشامل في شيلي، والذي من المقرر بدء عمله عام 2016، سيتمكن من جمع هذا الكم نفسه من البيانات كل خمسة أيام.

يمكن أيضًا إيجاد هذا الكم الهائل من البيانات بجانبنا. عندما بدأ العلماء في فك شفرات الحمض النووي البشري في عام 2003، تطلب منهم عقودًا من العمل المكثف من أجل ترتيب القاعدة المكونة من ثلاثة مليارات زوج من الأمشاج. والآن، بعد عقد كامل من الزمن، هناك منشأة واحدة يمكنها عمل ذلك في يوم واحد. في عالم المال، تتغير ملكية سبعة مليارات سهم كل يوم في سوق الأوراق المالية الأمريكية، وتحدث حوالي ثلثي تلك الصفقات عن طريق خوارزميات حاسوبية تعتمد على نماذج رياضية تستهلك جبالًا من البيانات لتوقع الأرباح ومحاولة تقليل المخاطرة.

لقد أغرقت شركات الإنترنت بكم هائل من البيانات؛ حيث تعالج شركة جوجل ما يزيد على 24 بيتابايت من البيانات كل يوم، أي ما يعادل جميع المطبوعات في مكتبة الكونجرس الأمريكي آلاف المرات. شركة فيس بوك التي لم تكن موجودة منذ عقد مضى، تُرفع على خوادمها ما يزيد على 10 ملايين صورة جديدة كل ساعة، ويضغط مستخدمو فيس بوك زر "أعجبني" أو يعلقون على الموضوعات حوالي 3 مليارات مرة كل يوم، تاركين خلفهم أثرًا رقميًّا يمكن للشركة فحصه لتتعلم المزيد عن مفضلات المستخدمين. في الوقت ذاته، يقوم عدد يقدر بثمانمائة مليون

مستخدم شهري لخدمة Youtube التابعة لشركة جوجل برفع ما يعادل ساعة من المقاطع المصورة كل ثانية. ويتزايد معدل الرسائل عبر موقع Twitter بمقدار 200% كل عام ووصل في عام 2012 إلى ما يزيد على 400 مليون تغريدة في اليوم.

من العلم والرعاية الصحية وعالم البنوك إلى عالم الإنترنت، قد تتنوع المجالات إلا أنها تشترك جميعًا فيما يلي: أن كم البيانات في العالم يتزايد بسرعة كبيرة، بصورة تفوق ليس فقط آلاتنا بل أيضًا تخيلاتنا.

حاول الكثيرون وضع تصور محدد لكمية المعلومات التي تحيط بنا وحساب السرعة التي تنمو بها، وحققوا درجات مختلفة من النجاح لأنهم قاموا بقياس أمور مختلفة. في إحدى الدراسات الشاملة التي أجراها "مارتن هيلبرت" من جامعة كارولينا الجنوبية مدرسة "أنينبرج" للاتصالات والصحافة، كان يتوق لوضع تصور لكل ما تم إنتاجه وتخزينه والتناقش حوله؛ مما يعني أنه لن يدرج الكتب والمطبوعات ورسائل البريد الإلكتروني والصور ومقاطع الموسيقى ومقاطع الفيديو (التناظرية والرقمية) فحسب، بل أيضًا ألعاب الفيديو والمحادثات الهاتفية وحتى أنظمة الملاحة في السيارات والخطابات المرسلة عبر البريد. أدرج كذلك البث الإعلامي مثل التلفاز والإذاعة بناءً على ما يشاهده الجمهور.

طبقًا لتقدير "هيلبرت"، كان هناك ما يزيد على 300 إكسابايت من البيانات المخزنة عام 2007. وحتى نفهم ما يعنيه هذا الأمر بمصطلحات يمكننا فهمها، فكر في الأمر على النحو التالي، فيلم كامل في صورته الرقمية يمكن ضغطه في ملف يبلغ حجمه جيجابايت واحد فقط، والإكسابايت يساوي مليار جيجابايت. المثير للدهشة أنه في عام 2007 كان هناك حوالي 7 % فقط من البيانات في صورتها التناظرية (أوراق، كتب، مطبوعات مصورة وغيرها)، وكانت البقية بيانات رقمية. ولكن لم يكن الوضع مختلفًا منذ وقت قريب. رغم أن أفكار "ثورة المعلومات" و"العصر الرقمي" كانت موجودة منذ ستينيات القرن العشرين، فإنها لم تتحقق إلا بصورة طفيفة. بحلول عام 2000، كان ربع المعلومات مخزنًا فقط في الصورة الرقمية، وكانت الأرباع الثلاثة الباقية في صورة أوراق وأفلام وأسطوانات الفينيل وشرائط الكاسيت وغيرها.

في حينها كان كم المعلومات الرقمية ليس كبيرًا ــ فكرة مهينة لمن كان يتصفح شبكة الإنترنت ويشتري الكتب عبر الإنترنت منذ فترة طويلة. (في الحقيقة، في عام 1986، كان حوالي 40 % من القوة الحاسوبية ذات الأغراض العامة في العالم تتمثل في الآلات الحاسبة بحجم الجيب والتي كانت تمتلك قوة معالجة أكبر من جميع أجهزة الحاسب الشخصية في ذلك الوقت). ولكن لأن البيانات الرقمية كانت تتوسع بسرعة كبيرة ــ حيث كانت تتضاعف كل ثلاث سنوات أو أكثر قليلاً طبقًا لما قاله "هيلبرت" ــ فقد تغير الوضع بسرعة، بينما لم تكن المعلومات في الصورة التناظرية تتطور على الإطلاق؛ لذا بحلول عام 2013، ستقدر كمية المعلومات المخزنة في العالم بحوالي 1.200 إكسابايت، والمعلومات غير الرقمية منها ستقل عن 2 %.

لا توجد طريقة جيدة للتفكير فيما يعنيه حجم البيانات. لو طُبع هذا الكم من

المعلومات في كتب لغطت مساحة الولايات المتحدة الأمريكية بأكملها بسمك 52 طبقة. ولو خُزِّنت على أقراص مدمجة، لغطت المسافة بين الأرض والقمر متراكمة فوق بعضها في خمس كومات منفصلة. في القرن الثالث قبل الميلاد، كان بطليموس الثاني ملك مصر يتوق إلى تخزين كل الأعمال المكتوبة في العالم؛ لذا مثلت مكتبة الإسكندرية العظيمة مجموع كل المعارف في العالم في وقتها. إن الفيضان الرقمي حاليًّا الذي يغمر الكرة الأرضية يمنح لكل إنسان يحيا على وجه الأرض في الوقت الحالي كمًّا من المعلومات يزيد بمقدار 320 مرة على كم المعلومات الذي يُقدر أنه كان مخزنًا في مكتبة الإسكندرية.

إن الأمور تتسارع من حولنا بدرجة كبيرة؛ حيث إن كمية المعلومات المخزنة تتزايد بمعدل يصل إلى أربعة أضعاف نمو الاقتصاد العالمي، في حين تنمو قوة المعالجة الحاسوبية بمقدار تسع مرات أسرع من الاقتصاد العالمي. أنا لا أتعجب من شكوى الناس من التحميل الزائد للمعلومات؛ حيث إن الجميع قد أصبحوا مقيدين بالتغيرات التي حدثت.

فكر في تأثير هذه الأمور على المستقبل من خلال المقارنة بين فيضان البيانات الحالي وثورة المعلومات السابقة التي تزامنت مع اختراع "جوتنبرج" للآلة الطباعة عام 1439؛ حيث طُبع في الخمسين سنة ما بين عامي 1453 و 1503 حوالي ثمانية ملايين كتاب طبقًا للمؤرخة "إليزابيث أينشتاين". ويزيد هذا العدد على جميع المخطوطات التي أنتجتها أوربا منذ إنشاء القسطنطينية قبل 1200 عام. بمعنى آخر، استغرق الأمر 50 عامًا حتى يتضاعف مخزون المعلومات في أوربا مقارنة بتضاعفه في الوقت الحالي كل ثلاث سنوات.

ما الذي تعنيه هذه الزيادة؟ يحب "بيتر نورفيج"، أحد خبراء الذكاء الاصطناعي في شركة جوجل، أن يفكر في الأمر بتشبيه تصوري. في البداية، يطلب منا أن نتأمل الحصان المميز المرسوم على جدار أحد الكهوف في لاسكو بفرنسا، والذي تشير الدراسات إلى أنه رُسم منذ حوالي 17 ألف سنة في العصر الحجري. ثم تأمل صورة ضوئية لحصان, أو لأفضل رسومات بابلو بيكاسو التي لا تختلف عن الرسومات في الكهف. في الحقيقة، عندما رأى "بيكاسو" رسومات لاسكو قال ساخرًا: "إننا لم نبتكر شيئًا جديدًا".

كانت كلمات "بيكاسو" صحيحة في أحد جوانبها ولكنه خالفه الصواب في جانب آخر. تذكر الصورة الضوئية للحصان. كان الأمر يستغرق وقتًا طويلاً فيما مضى لرسم صورة لحصان في حين أصبح في الوقت الحالي يستغرق وقتًا أقل بكثير بفضل التصوير الفوتوغرافي. هذا يعتبر تغييرًا، ولكنه لا يعتبر تغييرًا جذريًّا: حيث إنه في نهاية الأمر يعطي صورة للحصان. والآن، يطلب "نورفيج" أن تلتقط صورة للحصان وتسرعها بمقدار 24 صورة في الثانية. الآن أنتج التغيير الكمي تغيرًا نوعيًّا؛ حيث إن الفيلم يختلف جذريًّا عن الصورة الثابتة للحصان. ينطبق الأمر ذاته على البيانات الضخمة: من خلال تغيير الكمية، نغير الجوهر.

تأمل التشبيه التالي من مجال التكنولوجيا الدقيقة ــ حيث تصبح الأشياء أصغر

حجمًا لا أكبر حجمًا. المبدأ الذي تقوم عليه النانو تكنولوجي هو أنه عندما تصل إلى المستوى دون الذري يمكن للخواص المادية أن تتغير. بمعرفة هذه السمات الجديدة يمكنك أن تبتكر أشياء لتفعلها المادة لم تكن لتفعلها من قبل. على مستوى النانو، وعلى سبيل المثال، يمكن إنشاء معادن أكثر ليونة وسيراميك قابل للمد. والعكس صحيح، عندما نزيد كم البيانات التي نعمل بها، سيمكننا أن نفعل أمورًا لم يكن من الممكن فعلها عندما كنا نعمل على الكميات الصغيرة.

أحيانًا ما تكون المقتضيات التي نحيا بها ونزعم أنها ثابتة في كل الأمور، في حقيقتها مجرد وظائف للمقياس الذي نعمل من خلاله. هناك تشبيه ثالث، من العلوم أيضًا. بالنسبة للبشر، قانون الجاذبية هو أهم القوانين الفيزيائية في حياتنا: حيث يتحكم في كل ما نفعله. ولكن الجاذبية بالنسبة للحشرات الصغيرة لا تعد شيئًا مهمًّا على الإطلاق. بالنسبة لبعض الحشرات، مثل حشرة الماء، القانون الأهم في عالم الفيزياء هو قانون التوتر السطحي الذي يسمح لها بأن تسير على سطح الماء دون الوقوع فيه.

أما المعلومات، فهي مثل الفيزياء، يكون الحجم فيها مهمًّا. لذا كانت شركة جوجل قادرة على تحديد تفشي مرض الإنفلونزا بالدقة نفسها تقريبًا مثلما فعلت البيانات الرسمية المعتمدة على زيارات المرضى الحقيقية للأطباء؛ حيث أمكنها القيام بهذا من خلال الفحص المدقق لمئات المليارات من مصطلحات البحث ــ ويمكنها أن تحصل على نتيجة في وقت قريب، أسرع بكثير من نتائج المصادر الرسمية. وبالمثل، يمكن لنظام "فايركاست" الذي أنشأه "إتزيوني" بتوقع تباين أسعار تذاكر الطيران وبالتالي يقوم بتحويل قوة اقتصادية مهمة إلى أيدي المستهلكين. ولكن يمكن للنظامين أن يعملا بصورة جيدة فقط إذا ما توافرت لهما مئات المليارات من البيانات ليقوما بتحليلها.

المثالان السابقان يؤكدان الأهمية العلمية والاجتماعية للبيانات الضخمة وكذلك الدرجة التي يمكن أن تصل إليها البيانات الضخمة كمصدر للقيمة الاقتصادية؛ حيث يُظهران طريقتين لما يمكن للبيانات الضخمة فعله لزلزلة كل شيء من الشركات التجارية والعلوم إلى الرعاية الصحية والحكومات والتعليم والاقتصاد والعلوم الإنسانية وكل من مجالات المجتمع الأخرى.

رغم أن البيانات الضخمة ما زالت في بدايتها، فإننا نعتمد عليها بشكل يومي؛ حيث صُممت مرشحات البريد الإلكتروني العشوائي لكي تتكيف تلقائيًّا مع تغير الرسائل الإلكترونية غير المرغوب فيها: لا يمكن أن تتم برمجة البرنامج ليعرف كيفية منع العدد اللانهائي من الكلمات ذات الهجاء الخاطئ. تقوم مواقع التعارف بالجمع بين شخصين بناءً على الكثير من الصفات المشتركة بينهما ومقارنتها باختيارات نجحت فيما مضى. خاصية "التصحيح التلقائي" في الهواتف الذكية تتبع ما نفعله وتضيف كلمات جديدة إلى قاموسها بناءً على ما نكتب. ولكن ما زالت هذه الاستخدامات مجرد البداية. بداية من السيارات التي تحدد متى تنحرف ومتى تتوقف إلى حاسب "واتسون" من شركة آي بي إم القادر على هزيمة البشر في برنامج المسابقات "**جيوباردي**"، من شأن هذه المقاربة أن تغير الكثير من جوانب العالم الذي نحيا

به.

إن البيانات الضخمة في جوهرها تقوم على التوقع. رغم أنها تتبع فرعًا من علوم الحاسب يُطلق عليه الذكاء الاصطناعي، أو بالأحرى مجالًا يُدعى تعليم الآلات، إلا أن هذا التصوير يكون مضللاً. إن البيانات الضخمة ليست أن تحاول "تعليم" جهاز الحاسب أن يفكر" مثل البشر، بل هي تطبيق الرياضيات على كميات كبيرة من البيانات لكي نتمكن من استنتاج الاحتمالات: احتمالية أن تكون رسالة بريد إلكتروني ما من ضمن رسائل البريد العشوائي، أو احتمالية أن حروف "يناب" تعني كلمة "بيان"، أو أن مسار شخص ما وسرعته عبر الطريق بشكل خاطئ سيمكنانه من عبور الطريق في الوقت المناسب ــ على السيارة التي تقود نفسها أن تبطئ قليلاً. إن هذه الأنظمة تعمل على الوجه المثل لأنها تُغذى بالكثير من البيانات التي تبني عليها توقعاتها. علاوة على ذلك، بُنيت الأنظمة لتحسين نفسها طوال الوقت عن طريق متابعة أفضل المؤشرات والأنماط لتبحث عنها عند الحصول على المزيد من البيانات.

في المستقبل ــ وأقرب مما نظن ــ سيتم تحسين الكثير من أوجه عالمنا أو تُستبدل بها أنظمة الحاسب والتي تخضع في الوقت الحالي للحكم البشري وحده. ليس فقط قيادة السيارات أو المواءمة بين الأمور المختلفة، بل في أمور أكثر تعقيدًا. فرغم كل شيء، يمكن لموقع أمازون أن يقترح عليك الكتاب المناسب، ويمكن لمحرك بحث جوجل أن يرتب لك أكثر موقع متعلق بما تبحث عنه، ويعرف موقع فيس بوك ما يعجبك، ويتكهن موقع لينكد إن بمن نعرفهم. سيتم تطبيق تلك التقنيات نفسها من أجل تشخيص الأمراض وتحديد الأدوية المعالجة لها، وربما تحديد "المجرمين" قبل ارتكابهم الجرم. كما تمكنت شبكة الإنترنت من تغيير العالم بشكل جذري من خلال إضافة الاتصالات إلى الحاسب، ستقوم أيضًا البيانات الضخمة بتغيير أوجه معينة من حياتنا من خلال إعطائها بعدًا كميًّا لم يُعْطَ لها من قبل.

المزيد، الفوضوي، الجيد كُلٌّ بدرجة كافية

ستكون البيانات الضخمة مصدرًا للقيم الاقتصادية والابتكارية الجديدة، بل وستكون مصدرًا لأكثر من ذلك. إن هيمنة البيانات الضخمة تمثل ثلاثة تحولات في الطرق التي نحلل بها المعلومات من شأنها تغيير الطريقة التي نستوعب وننظم بها المجتمع.

التحول الأول سنتناوله في الفصل الثاني. في هذا العالم الجديد سيمكننا أن نحلل أمورًا أكثر من البيانات؛ ففي بعض الحالات سيمكننا أن نعالج **جميع** الأمور المتعلقة بظاهرة بعينها. منذ القرن التاسع عشر، كان المجتمع يعتمد على استخدام العينات عندما كانت تواجهه أعداد كبيرة مما يقوم بتحليله، إلا أن الحاجة للحصول على عينات أصبحت من الآثار الغابرة لزمن كانت فيه المعلومات نادرة، وكانت نتاج القيود الطبيعية للتعامل مع المعلومات في هذه الحقبة التناظرية. قبل هيمنة التقنيات الرقمية عالية الأداء، لم نكن ندرك أن أخذ عينات مما كنا نقوم بتحليله

يعتبر من القيود غير الطبيعية ـــ حيث إننا عادة ما كنا نأخذ الأمر على أنه مسلم به. إن استخدام البيانات جميعها يساعدنا على أن نرى التفاصيل التي لم نكن لنراها إذا ما كنا مقيدين بالكميات المحدودة من المعلومات. إن البيانات الضخمة تمنحنا رؤية واضحة للمستوى الأدنى: الفئات الثانوية والأسواق الثانوية التي لا يمكن للعينات تقييمها.

إن إلقاء الضوء على المزيد من البيانات يسمح لنا أيضًا بأن نطلق العنان لرغبتنا في تحري الدقة، وهذا هو التحول الثاني الذي سنتناوله في الفصل الثالث. إن الأمر عبارة عن علاقة تبادلية: عندما تقل الأخطاء الناتجة عن أخذ عينات من البيانات يمكننا أن نتقبل المزيد من أخطاء القياس. عندما تكون قدرتنا على القياس محدودة، يمكننا فقط أن نحسب الأمور ذات الأهمية القصوى. إن العمل على الحصول على الأعداد الدقيقة أمر مناسب تمامًا؛ حيث إنه لا فائدة من بيع المواشي إذا لم يكن البائع يعلم ما إذا كان هناك 100 بقرة أم 80 فقط في القطيع. حتى وقت قريب، كانت جميع أدواتنا الرقمية قائمة على تحري الدقة العددية: حيث افترضنا أن محركات قواعد البيانات قد تستخرج السجلات التي تناسب بدقة طلبات البحث التي نكتبها، وتضعها فيما يشبه الجداول الممتدة مصنفة الأرقام في أعمدة.

إن هذا النوع من التفكير يعتبر من سمات بيئة "البيانات الصغيرة": مع وجود عدد قليل من الأمور لقياسها، وكان علينا أن نبسط ما قمنا بتعقيده حتى نتمكن من حسابه بأكبر قدر ممكن من الدقة. قد يكون هذا الأمر واضحًا بصورة ما: يمكن لمتجر صغير أن يعد النقود في ماكينة الصراف في نهاية اليوم حتى آخر سنت، ولكن لا يمكننا فعل الأمر ذاته مع إجمالي الناتج المحلي لدولة ما. مع زيادة المقياس، تزداد عدم الدقة بدورها.

تحتاج الدقة إلى بيانات دقيقة للغاية، وقد تصلح للكميات الصغيرة من البيانات، ولا شك في أن بعض المواقف ما زالت تحتاج إلى هذا الأسلوب: سواء كان المرء يمتلك مبالغ كافية من المال في المصرف أو لا يمتلكها من أجل كتابة شيك لشخص آخر. ولكن في المقابل ومن أجل استخدام مجموعات أكثر شمولاً من البيانات يمكننا أن نتخلى عن بعض من الدقة الصارمة في عالم البيانات الضخمة.

عادة ما تكون البيانات الضخمة فوضوية ومتباينة الجودة وموزعة على عدد لا يحصى من الخوادم في جميع أنحاء العالم. مع البيانات الضخمة، سنرتضي دائمًا بحس التوجه العام بدلاً من التعرف على ظاهرة ما بأدق تفاصيلها. إننا لن نتخلى عن الدقة تمامًا، بل سنتخلى عن تفانينا في تحقيقها؛ حيث إن ما نفقده من الدقة على المستوى الصغير نجني من ورائه أفكارًا عميقة على المستوى الأكبر.

هذان التحولان يقودانا إلى تغيير ثالث، والذي سنتناوله في الفصل الرابع: الابتعاد عن البحث العتيق عن السببية. إننا كبشر نسعى دائمًا للبحث عن الأسباب، حتى إن كان البحث عن السببية صعبًا ويجعلنا نضل الطريق. أما في عالم البيانات الضخمة، فعلى النقيض، ليس علينا أن نتقيد بالأسباب، بل يمكننا أن نكتشف بدلاً منها أنماطًا وعلاقات تبادلية داخل البيانات تقدم لنا أفكارًا حديثة وقيمة. قد لا تخبرنا العلاقات التبادلية بسبب حدوث هذا الأمر، ولكنها تنبهنا إلى أن **الأمر** يحدث في الوقت

الحالي.
ويعتبر هذا الأمر في الكثير من المواقف جيدًا بدرجة كافية. إذا كانت الملايين من التقارير الطبية الإلكترونية تكشف عن أن من يعانون مرض السرطان الذين تناولوا نوعًا معينًا من الأسبرين مع عصير البرتقال قد عولجوا من المرض، فإن السبب الرئيسي لتحسن صحتهم قد يكون أقل أهمية من حقيقة أنهم سينجون من المرض. وبالمثل، إذا تمكنا من توفير المال من خلال إدراكنا أفضل وقت لشراء تذاكر الطيران دون معرفة السبب في الجنون الذي يصيب أسعار التذاكر، فإنه أمر جيد. إن البيانات الضخمة تدور حول "**الأمر ذاته**" لا "**سببه**". لسنا دائمًا بحاجة لمعرفة السبب وراء الظاهرة، بل يمكننا أن نترك البيانات تتحدث عن نفسها.
قبل البيانات الضخمة، كانت تحليلاتنا قاصرة على اختبار عدد محدود من الفرضيات التي وضعناها جيدًا قبل أن نجمع البيانات. عندما نترك البيانات لتتحدث، سيمكننا أن نحصل على علاقات لم نتصور وجودها؛ لذا تعمل بعض التمويلات الخفية على حمل موقع تويتر على التنبؤ بأداء سوق الأوراق المالية، وتبني مواقع أمازون ونتفليكس اقتراحات المنتجات التي تقدمها للمستهلكين على أساس عشرات الآلاف من مشاركات المستخدمين في تلك المواقع، وتقوم مواقع تويتر ولينكد إن وفيس بوك بعمل "رسم بياني اجتماعي" لعلاقات المستخدمين لتتمكن من معرفة ما يفضلون.

لا شك في أن البشر كانوا يقومون بتحليل البيانات منذ قرن مضي؛ حيث اخترعت الكتابة في بلاد ما بين الرافدين لأن المسئولين كانوا يرغبون في أداة فعالة لتسجيل المعلومات وحفظها. منذ العصور القديمة، كانت الحكومات تقوم بعمل إحصاء لعدد السكان لتجمع كم هائل من البيانات عن شعوبها، وعلى مدى مائتي عام، كان المسئولون عن شئون التأمين يقومون بالمثل من أجل جمع كم كبير من البيانات بخصوص المخاطر التي يرغبون في فهمها ـــ أو على الأقل تفاديها.
ولكن كان جمع تلك البيانات في العصر التناظري وتحليلها مكلفًا بدرجة كبيرة ويستهلك الكثير من الوقت. وكانت المسائل الجديدة تعني أن البيانات يجب أن يتم جمعها من جديد وأن تتم التحليلات منذ البداية.
جاءت الخطوة الكبيرة نحو إدارة البيانات بصورة أكثر فاعلية مع قدوم الرقمنة: تمكين الحواسب من قراءة المعلومات التناظرية، الأمر الذي جعل تخزين البيانات ومعالجتها أسهل وأرخص؛ حيث حسن هذا التطوير من فاعلية معالجة البيانات بشكل جذري. إن جمع البيانات وتحليلها اللذين كانا يستغرقان فيما مضى أعوامًا كاملة يحدث الآن في غضون أيام وربما أقل. ولكن الأمر الذي لم يتغير كثيرًا، هو أن الأشخاص القائمين على تحليل البيانات كانوا منغمسين بدرجة كبيرة في النموذج الفكري من العصر التناظري الذي يزعم أن البيانات وحيدة الغرض وتنتهي قيمتها بانتهاء الغرض منها، وساعدت عملياتنا تلك هذا الإجحاف على أن يستمر. الأمر الذي يحمل درجة الأهمية نفسها هو أن الرقمنة قد وجدت لتمكين التحول إلى استخدام البيانات الضخمة، في حين أن وجود أجهزة الحاسب في حد ذاته لم يكن

هو القوة الدافعة للبيانات الضخمة.

لا يوجد مصطلح يصلح لوصف ما يحدث الآن، ولكن يوجد مصطلح واحد من شأنه المساعدة على تشكيل التغيير وهو **سيطرة البيانات**، المفهوم الذي سنتناوله في الفصل الخامس. يشير المفهوم إلى الحصول على معلومات عن كل ما يوجد على سطح الأرض ـــ من بينها أمور لم نفكر فيها قط على أنها معلومات، مثل مكان وجود شخص ما، والترددات التي تصدرها المحركات، أو قدر الضغط على جسر ما ـــ ومن ثم تحويلها إلى بيانات ليتم حسابها؛ مما يسمح لنا باستخدام المعلومات بطرق جديدة، مثل استخدامها في التحليل التوقعي: اكتشاف ما إذا كان المحرك سيتعطل بناءً على الحرارة والترددات التي يصدرها. ونتيجة لهذا، يمكننا اكتشاف القيمة الضمنية والمختفية للمعلومات.

إننا بصدد القيام برحلة لاقتناص الكنوز، مدفوعة بالأفكار المراد استخراجها من البيانات والقيمة الساكنة التي يمكن إطلاق سراحها من خلال التحول من السببية إلى العلاقات التبادلية. ولكن لا يوجد كنز واحد فحسب؛ حيث إن كل مجموعة من البيانات تحتوي على بعض القيم غير الملموسة والمختفية والتي لم تطفُ على السطح بعد، وقد بدأ السباق لاقتناص جميع هذه القيم واكتشافها.

إن البيانات الضخمة من شأنها تغيير طبيعة عالم الأعمال والأسواق والمجتمع ككل كما سنتناول في الفصلين السادس والسابع. في القرن العشرين، تحولت القيمة من البنية التحتية المادية مثل الأراضي والمصانع إلى أمور غير مادية مثل العلامات التجارية والملكية الفكرية، والتي تتسع الآن لتشمل البيانات، والتي تتحول بالتدريج لتصبح من الأصول الثابتة المهمة في المؤسسات، ومدخلًا اقتصاديًّا حيويًّا، وأساسًا لنماذج تجارية جديدة. إنها وقود اقتصاد المعلومات. رغم أن البيانات نادرًا ما تظهر في تقارير الموازنة للمؤسسات، فإن هذا الأمر سيتغير بمرور الوقت.

رغم وجود بعض التقنيات القائمة على معالجة البيانات منذ فترة، فإنها كانت تتوافر في الماضي للوكالات الاستخباراتية ومعامل الأبحاث والشركات العالمية الكبرى فقط. إلا أن شركتي وولمارت وكابيتال وان قادتا استخدام البيانات الضخمة في مجالات البيع بالتجزئة والمصارف مما مكنهما من تغيير أسلوب عمل هذين المجالين. في الوقت الحالي تحول عدد كبير من هذه الأدوات ليكون متاحًا على نطاق أوسع (رغم أن البيانات ذاتها لم تصبح كذلك).

لقد سبب تأثير البيانات على الأفراد صدمة كبرى للجميع. هناك مجالات معينة لا تحظى بأهمية كبيرة في عالم تكون فيه الاحتمالات والعلاقات المتبادلة على القمة. في فيلم *Moneyball*، أبعد الإحصائيون مكتشفي مواهب البيسبول عن صدارة المشهد عندما أفسحت الموهبة الفطرية المجال أمام التحليلات الدقيقة، وبالمثل، لن يبتعد المتخصصون في موضوع نقاشنا كثيرًا عن فعل المثل، ولكن ستكون عليهم منافسة ما تقوله تحليلات البيانات الضخمة؛ مما سيعمل على فرض تعديل على الأفكار التقليدية للإدارة واتخاذ القرار والموارد البشرية والتعليم.

إن أغلب مؤسساتنا قامت بناءً على فرضية أن القرارات البشرية تعتمد على قدر

يسير ودقيق وسببي بطبيعته من المعلومات. ولكن يختلف الأمر في حالة البيانات الضخمة؛ حيث يمكن معالجتها بسرعة مع عدم الإصرار على الدقة المتناهية للنتائج. علاوة على ذلك، بفضل الكم الهائل للبيانات، فإن القرارات عادة ما سيتم اتخاذها عن طريق الآلات لا البشر. وسنتأمل الجانب المظلم من البيانات الضخمة في الفصل الثامن.

إن المجتمعات تمتلك خبرة تمتد قرونًا فيما يتعلق بفهم السلوك البشري ومراقبته. ولكن كيف يمكن وضع نظام حسابي لحساب هذا الأمر؟ في الأيام الأولى للأنظمة الحاسوبية، أدرك واضعو السياسات كيف يمكن استخدام التكنولوجيا لتقويض الخصوصية. ومنذ ذلك الحين سنت المجتمعات قوانين لحماية المعلومات الشخصية. ولكن في عالم البيانات الضخمة، تشكل هذه القوانين خطًّا دفاعيًّا واهيًا يشبه خط ماجينو في فرنسا. إن الناس ينشرون المعلومات على شبكة الإنترنت عن رضا كامل منهم — وهذه سمة رئيسية لمثل هذه الخدمات، وبالتالي ليست بهذا الوهن لتتم حمايتها.

في الوقت ذاته، إن الخطر الكامن في تحول الأفراد من الخصوصية إلى الاحتمالية: ستتنبأ الأنظمة الحسابية باحتمالية إصابة شخص ما بنوبة قلبية (ويدفع المزيد إلى التأمين الصحي)، أو شخص آخر سيرتكب جريمة (وربما يتم القبض عليه قبل ارتكابه إياها)؛ مما يقودنا إلى اعتبار أخلاقي لدور الإرادة الحرة مقابل ديكتاتورية البيانات. هل تفضل الإرادة الفردية البيانات الضخمة، حتى وإن كانت الإحصائيات تقول العكس؟ كما مهدت الطباعة الطريق أمام القوانين التي تحمي حرية التعبير — والتي لم تكن موجودة من قبل بسبب أن النصوص التعبيرية المكتوبة كانت قليلة لدرجة عدم التفكير في حمايتها — فإن عصر البيانات الضخمة سيتطلب قوانين جديدة لحماية حرمة الأفراد.

بشتى الطرق، سيجب تغيير الطريقة التي نتحكم ونتعامل بها مع البيانات؛ حيث إننا على أعتاب عالم من التوقعات المستمرة التي تقودها البيانات حيث لن نتمكن من شرح الأسباب التي دفعتنا إلى اتخاذ القرارات. ماذا سيعني الأمر إذا لم يتمكن الطبيب من اتخاذ قرار بالتدخل الطبي قبل أن يطلب من المريض أن يستشير صندوقه الأسود، كما من المفترض أن يفعل الطبيب الذي يعتمد على التشخيص بواسطة البيانات الضخمة؟ هل ستحتاج معايير الأنظمة القضائية إلى تغيير مصطلح "السبب المحتمل" ليكون "سببية الاحتمال" — وإن كان الأمر على هذا المنوال، فما هو تأثير ذلك على حرية الإنسان وكرامته؟

يحتاج عصر البيانات الضخمة إلى ظهور مبادئ جديدة، والتي سنضعها معًا في الفصل التاسع. رغم أنها ستقوم على القيم التي تطورت وحُفظت في عالم البيانات الصغيرة، فإنها لا تعتبر مجرد إعادة إنعاش للقواعد القديمة لتطبيقها على الظروف الجديدة، بل إدراك الحاجة الملحة لوجود مبادئ جديدة تمامًا.

سيستفيد المجتمع من البيانات الضخمة بشكل هائل؛ حيث إن البيانات الضخمة ستصبح جزءًا من علاج المشكلات الملحة مثل تحديد التغيرات المناخية والقضاء على الأمراض وإنشاء حكومات جيدة والنمو الاقتصادي. ولكن يتحدانا عصر البيانات

الضخمة أيضًا لنصبح على استعداد أفضل للطرق التي من خلالها سيعمل ترويض التكنولوجيا على تغيير مؤسساتنا الاجتماعية وأنفسنا.

تعتبر البيانات الضخمة خطوة كبيرة في مسعى الجنس البشري لتحديد العالم من حولنا وفهمها؛ حيث إن كمًّا كبيرًا من الأشياء التي كان من المستحيل قياسها وتخزينها وتحليلها ونشرها من قبل سيتحول إلى صورة بيانات. إن ترويض كمٍّ كبير من البيانات بدلاً من كم ضئيل مع عدم تحري الدقة بشكل كبير من شأنه تمهيد الطريق أمام طرق جديدة للفهم تقود المجتمع للتخلي عن مفضلاته الوقتية للسببية، وفي كثير من الأحيان تجعل المجتمع يستفيد كثيرًا من العلاقات التبادلية.

النموذج المثالي المعبر عن تحديد آليات السببية هو وهم الإطراء على الذات الذي تقلبه البيانات الضخمة رأسًا على عقب. ولكننا وصلنا مرة أخرى إلى طريق مسدود "لا أمل منه". أي أن القناعات التي نؤمن بها في سبيلها للتغير مرة أخرى، ولكنها سيستبدل بها هذه المرة شيء أفضل. ما الدور المتبقي ليلعبه الحدس والإيمان وعدم الثقة والتصرف على العكس من الدليل والتعلم من خلال الخبرة؟ ولأن العالم يتحول من السببية إلى العلاقات التبادلية، فكيف يمكننا أن نتقدم للأمام دون أن نحط من قدر المؤسسات الاجتماعية والإنسانية والتقدم الذي اعتمد على المنطق؟ ينتوي هذا الكتاب أن يشرح لنا موقعنا الحالي من هذا الأمر، وأن يتتبع مسارنا وصولاً إلى هذا الموقع، ويقدم دليلاً نحتاج إليه للفوائد والمخاطر التي قد تواجهنا خلال رحلتنا القادمة.

2
المزيد

تتعلق البيانات الضخمة برؤية العلاقات واستيعابها داخل وبين أجزاء من المعلومات والتي ــ حتى وقت قريب ــ صارعنا كثيرًا حتى نتمكن من استيعابها بالكامل. يقول خبير البيانات الضخمة بشركة آي بي إم إنك بحاجة لأن تدع البيانات "تتحدث إليك". قد يبدو هذا الأمر مبتذلاً إلى حدٍّ ما. إن البشر كانوا ينظرون للبيانات ليتعلموا منها الكثير عن العالم منذ وقت طويل، سواء كان هذا الأمر من خلال الحس غير الرسمي النابع من عشرات الآلاف من الملاحظات التي نلاحظها كل يوم، أو، خصوصًا في القرنين الماضيين، من خلال الحس الرسمي بتحديد الوحدات التي يمكن التلاعب بها بواسطة النماذج الحسابية الفعالة.

ربما جعل العصر الرقمي من معالجة البيانات أمرًا أكثر سرعة وسهولة؛ حيث يمكننا حساب ملايين الأرقام في ثانية واحدة. ولكن عندما نتحدث عن البيانات التي تتحدث، فإننا نعني أمرًا أكثر أهمية ــ واختلافًا. كما أشرنا في الفصل الأول، تتعلق البيانات الضخمة بثلاثة تحولات رئيسية في العقليات المتصلة ببعضها وبالتالي تقوم بدعم بعضها الآخر. التحول الأول هو القدرة على تحليل كميات هائلة من البيانات حول موضوع بعينه بدلاً من التقيد بتحليل كميات قليلة منها. التحول الثاني هو الاستعداد لتقبل فوضى البيانات الضخمة بدلاً من التركيز على تحري الدقة الشديدة. التحول الثالث هو الاحترام الكبير للعلاقات التبادلية بدلاً من الاستمرار في السعي للبحث عن السببية المراوغة. يلقي هذا الفصل الضوء على التحول الأول: استخدام كمية البيانات التي نمتلكها كاملة بدلاً من استخدام كميات قليلة منها.

إن التحدي الكامن في معالجة الكميات الضخمة من البيانات بدقة يواجهنا منذ فترة؛ حيث إننا في أغلب مراحل تاريخنا كنا نعمل بكمية محدودة من البيانات لأن أدوات جمعها وتنظيمها وتخزينها وتحليلها كانت محدودة. كنا نقوم بتصفية المعلومات التي نعتمد عليها حتى أصغر جزء فيها لنتمكن من فحصها بسهولة أكبر، وقد كان هذا الأمر نوعًا من الرقابة الشخصية اللاواعية: تعاملنا مع صعوبة التفاعل مع البيانات على أنها واقع مؤسف، بدلاً من رؤيتها على حقيقتها؛ قيد اصطناعي وضعته التقنيات المستخدمة في ذلك الوقت. أما اليوم؛ فقد تحولت البيئة التقنية بمقدار 179 درجة، ولكن ما زالت هناك، وستظل دائمًا، قيودًا على كم المعلومات التي يمكننا إدارتها، ولكنها أقل محدودية بكثير من تلك التي استخدمت من قبل وستصبح أقل محدودية بمرور الزمن.

إننا لم نقدر حتى الآن بشكل كامل حريتنا المكتسبة حديثًا لجمع كميات أكبر من البيانات واستخدامها؛ حيث إن أغلب خبراتنا وتصميم مؤسساتنا قد افترض أن توافر المعلومات يجب أن يكون محدودًا. لقد اعتمدنا على قدرتنا على جمع كم محدود من المعلومات، وكان هذا الأمر هو ما قمنا به دومًا؛ بحيث أصبح هذا الأمر مرضيًا لنا

بالكامل. حتى إننا طورنا تقنيات متقنة لتستخدم أقل كم ممكن من البيانات؛ حيث كان أحد أهداف الإحصاء هو التأكد من الحصول على أفضل النتائج باستخدام أقل كم ممكن من البيانات. في الواقع، لقد نظمنا طرقنا لتقزيم كمية المعلومات التي استخدمناها في معاييرنا وعملياتنا وأنظمتنا التحفيزية. وحتى نتمكن من فهم ما يعنيه التحول لاستخدام البيانات الضخمة، تبدأ القصة بالعودة لإلقاء نظرة على الماضي.

حتى وقت قريب، لم تكن الشركات الخاصة، وحتى الأفراد في يومنا هذا، قادرة على جمع المعلومات وتصنيفها بكميات كبيرة. في الماضي، كانت هذه المهمة تقتصر على المؤسسات ذات النفوذ الكبير مثل دار العبادة والحكومة، الأمر الذي كان يحدث في كل المجتمعات على حدٍّ سواء. إن أقدم سجلات الحسابات كانت في حوالي عام 5000 قبل الميلاد؛ حيث كان التجار السومريون يستخدمون خرزات طينية صغيرة للإشارة إلى البضائع المعدة للتبادل.أما الحسابات الكبيرة فقد كانت من تخصص الحكومة. على مدار آلاف السنين، حاولت الحكومات أن تتبع شعوبها من خلال جمع المعلومات عنهم.

فيما يتعلق بالتعداد السكاني، يُقال إن المصريين القدماء قد استخدموا أسلوب التعداد السكاني وكذلك فعل الصينيون. وقد ذُكر التعداد السكاني في الكتب التاريخية القديمة؛ حيث ذُكر فيها أن التعداد السكاني قد أمر به القيصر "أوغستسوس" ـــ "يجب أن يدفع العالم بأكمله الضرائب". كتاب يوم الحساب في عام 1086، أحد أعظم الكنوز البريطانية، والذي لم يكن له مثيل في عصره؛ حيث كان عبارة عن تعداد شامل لكل الشعب البريطاني وأراضيهم وأملاكهم. كان أن انتشر المفوضون الملكيون في جميع القرى البريطانية لجمع المعلومات ووضعها في الكتاب ـــ والذي حمل فيما بعد اسم "يوم الحساب"؛ لأن الأسلوب الذي اتبع في جمع المعلومات كان يشبه يوم الحساب حينما تُكشف أعمال جميع البشر.

كان إجراء التعداد السكاني مكلفًا ويستغرق الكثير من الوقت، لدرجة أن الملك وليام الأول، الذي أمر بجمع المعلومات لكتاب يوم الحساب، لم يَعِشْ حتى يرى استكمال هذا الكتاب. ولكن كان البديل الوحيد لحمل هذا العبء هو التوقف عن جمع المعلومات. وبعد كل هذا الوقت والتكاليف، كانت المعلومات المجمعة تقريبية؛ حيث إن القائمين بالتعداد لم يتمكنوا من تعداد الجميع بدقة. كلمة تعداد ذاتها في الإنجليزية "Census" مشتقة من الكلمة اللاتينية "Censere" والتي تعني "التقدير".

منذ أكثر من ثلاثمائة سنة مضت، خطرت لبائع خردوات بريطاني يُدعى "جون جراونت" فكرة جديدة؛ حيث كان يرغب في معرفة عدد سكان لندن خلال وقت الوباء. وبدلاً من أن يقوم بتعداد الجميع، ابتدع أسلوبًا جديدًا ـــ الأمر الذي نطلق عليه في الوقت الحالي "إحصائيات" ـــ مكنه من **تخمين** عدد السكان. كان أسلوبه بدائيًّا، ولكنه كان نواة الفكرة بأن المرء قادر على أن يُكَوِّن ـــ استقرائيًّا ـــ من عينة صغيرة فكرة جيدة عن عدد السكان الإجمالي. ولكن كيف يمكن للمرء فعل هذا؟ ذلك هو الأمر المهم،؛ حيث قام "جراونت" باستخدام عينته لقياس حجم

السكان ككل.

وقد اشتهر أسلوبه كثيرًا، رغم أننا علمنا فيما بعد أن هذه الأرقام المنطقية التي توصل إليها، قد كانت مصادفة. وعلى مر الأجيال، استمرت العيوب في الظهور في أسلوب العينات. فبالنسبة للتعداد السكاني وغيره من المشروعات المتعلقة بـ "بيانات ضخمة"، حاول الناس تطبيق هذا الأسلوب القوي المتعلق بحساب كل رقم نستخدمه في حياتنا.

بسبب أن التعدادات السكانية كانت غاية في التعقيد ومكلفة للغاية وتستغرق الكثير من الوقت، كانت نادرًا ما تُجرى؛ حيث إن الرومان القدامى، والذين كان يتعدى تعداد شعوبهم مئات الألوف، كانوا يجرون التعداد السكاني مرة كل خمس سنوات. ونص الدستور الأمريكي على أن يُجرى التعداد السكاني مرة كل عقد؛ حيث إن تلك الدولة النامية في حينها كانت تقدر تعدادها السكاني بالملايين. ولكن في أواخر القرن التاسع عشر كان هذا الأمر يشكل مشكلة كبيرة؛ حيث إن البيانات التي كان يتم تجميعها كانت تفوق قدرة مكتب التعداد السكاني على مجاراتها.

استغرق التعداد السكاني عام 1880 ثماني سنوات لاستكماله، وكانت المعلومات المجمعة قد عفا عليها الزمن قبل حتى أن تصبح متداولة. الأمر الأسوأ، أن المسئولين قدروا أن التعداد السكاني لعام 1890 سيستغرق 13 عامًا كاملاً ليتم تنظيمه في جداول ـــ يا له من أمر مثير للسخرية، ناهيك عن عدم التقيد بما نص عليه الدستور. ولأن تحديد الضرائب وعضوية الكونجرس كانت تعتمد على عدد السكان، فإن الحصول على تعداد سكاني صحيح بل وفوري كان أمرًا ضروريًّا.

كانت المشكلة التي واجهت مكتب التعداد السكاني الأمريكي شبيهة بالصعوبة التي واجهها العلماء ورجال الأعمال في بداية الألفية الجديدة، عندما تبين لهم أنهم يغرقون في كم هائل من البيانات: كانت كمية المعلومات التي جُمعت تفوق بكثير قدرة الأدوات المستخدمة في معالجتها، وكانوا بحاجة إلى تقنية جديدة لتحليل تلك البيانات. في ثمانينيات القرن التاسع عشر، كان الوضع ميئوسًا منه مما دفع مكتب التعداد السكاني الأمريكي للتعاقد مع المخترع الأمريكي "هيرمان هوليريث" لاستخدام أفكاره عن البطاقات المخرمة وآلات الجدولة في التعداد السكاني لعام 1890.

وبجهد كبير، تمكن "هيرمان" من تقليص فترة التعداد السكاني من ثماني سنوات إلى ما يقل عن العام الواحد. وقد كان هذا العمل من الأعمال الفذة التي أعلنت عن بداية أسلوب معالجة البيانات الآلي (والذي وضع حجر الأساس لما أصبح فيما بعد شركة آي بي إم). ولكن كان هذا الأسلوب لجمع البيانات الضخمة وتحليلها لا يزال باهظًا للغاية، فقد كان على كل فرد من مواطني الولايات المتحدة الأمريكية أن يملأ استمارة بمعلوماته الشخصية وأن يتم نقل البيانات على بطاقات مخرمة والتي كانت تُستخدم بدورها في آلات الجدولة. مع وجود مثل هذا الأسلوب المكلف، كان من الصعب التخيل أن يتم تكرار إجراء هذا التعداد السكاني في فترة بينية تقل عن عقد كامل، رغم أن هذا التأخير لم يكن مفيدًا لدولة تنمو بسرعة كبيرة.

وهنا يكمن لب الأمر: استخدام كل البيانات المتاحة، أم القليل منها فقط؟ إن

الحصول على جميع البيانات حول الأمر المراد قياسه لا شك في أنه الأسلوب المنطقي للعمل، ولكنه لا يكون عمليًّا طوال الوقت عندما يكون كم المعلومات كبيرًا. ولكن كيف يمكننا اختيار عينة بعينها؟ قال بعض الناس إن إنشاء عينة معبرة عن الكم الإجمالي للبيانات ربما تكون الطريقة المناسبة للتقدم. ولكن في عام 19، أظهر الإحصائي البولندي "جيرزي نيمان" أن مثل هذه الطريقة تؤدي إلى أخطاء كبيرة، وأن الطريقة المناسبة لفعل ذلك الأمر هي اختيار عينة عشوائية.

وضح الإحصائيون أن دقة العينات تتزايد بشكل كبير عندما تكون العينات عشوائية، وليس بزيادة حجمها. في الحقيقة، ورغم أن هذا الأمر قد يبدو مدهشًا، فإن عينة مختارة عشوائيًّا مكونة من إجابات 1100 شخص على سؤال ثنائي (نعم أو لا، مع وجود فرضيات متساوية للإجابتين) تكون معبرة بشكل مذهل عن عدد السكان بأكمله. كان هذا في 19 حالة من مجمل 20 حالة مع وجود هامش خطأ يبلغ 3%، بغض النظر عما إذا كان عدد السكان يبلغ مائة ألف أو مائة مليون. لماذا يجب أن يكون الأمر معقدًا رياضيًّا بهذا الشكل، ولكن الإجابة المختصرة هي أنه في مرحلة معينة في المستقبل القريب، مع تزايد حجم الأرقام أكثر فأكثر، فإن الكمية الهامشية للمعلومات الجديدة التي نتعلمها من كل ملاحظة تقل أكثر فأكثر.

حقيقة أن العشوائية تغلبت على حجم العينة كانت فكرة مدهشة؛ حيث إنها مهدت الطريق أمام الطرق الجديدة لجمع المعلومات لتزدهر. إن البيانات التي تستخدم العينات العشوائية يمكن جمعها بتكلفة زهيدة ويمكن من خلالها تقدير العدد الإجمالي استقرائيًّا بدقة متناهية. نتيجة لهذا الأمر، يمكن للحكومات أن تُجري نسخًا صغيرة من التعدادات السكانية باستخدام العينات العشوائية كل عام، بدلاً من مرة كل عقد. وقد فعلت الحكومات ذلك بالفعل؛ فعلى سبيل المثال، المكتب الأمريكي للتعداد السكاني يُجري ما يزيد على مائتي استقصاء رأي اقتصادي وديموغرافي كل عام باستخدام العينات العشوائية، إلى جانب التعداد السكاني الذي يجريه كل عشر سنوات الذي يشمل جميع المواطنين الأمريكيين. كانت العينات العشوائية حلاً لمشكلة زيادة تحميل المعلومات في العصور القديمة عندما كان جمع البيانات وتحليلها أمرًا بالغ الصعوبة.

سرعان ما تعدى تطبيق هذه الطريقة مجرد استخدامها في القطاع العام والتعداد السكاني، ولأن العينات العشوائية في جوهرها تقلل من مشكلات البيانات الضخمة وتحولها إلى مشكلات بيانية يمكن إدارتها، فقد تم استخدامها في عالم الأعمال من أجل التأكد من جودة التصنيع ـــ إدخال التحسينات بتكاليف أقل وبطرق أسهل. إن إدارة الجودة الشاملة تحتاج إلى فحص كل منتج يخرج من السير المتحرك بالمصانع، ولكن الآن تعتبر عينة عشوائية من بين مجموعة من المنتجات ويتم اختبارها أمرًا كافيًا. وبالمثل، تم استخدام الطريقة ذاتها في استقصاءات رأي المستهلكين في مجال البيع بالتجزئة والاقتراعات المفاجئة في عالم السياسة. وقد حولت جزءًا كبيرًا مما كنا ندعوه بالعلوم الإنسانية ليصبح **علوم** المجتمع.

لقد حظيت العينات العشوائية بنجاح باهر وتعتبر حجر الأساس للقياسات الحديثة على نطاق واسع. ولكنها ليست سوى طريق مختصر، بديل ممتاز آخر لجمع

مجموعات كاملة من البيانات وتحليلها. وهي تتسم بعدد من نقاط الضعف الفطرية؛ حيث إن دقتها تعتمد على عشوائية اختيار بيانات العينة، ولكن تحقيق مثل هذه العشوائية قد يكون خادعًا؛ حيث إن التوجهات المنهجية في أسلوب جمع البيانات قد تؤدي إلى نتائج اعتبارية خاطئة.

هناك أصداء لمثل هذه المشكلات في الاقتراع للانتخابات بواسطة خطوط الهاتف الأرضية؛ فالعينة تعد متحيزة حيث تجاهلت العينة العشوائية الأشخاص الذين يستخدمون الهواتف النقالة فقط (أشخاص أصغر سنًّا وأكثر تحررًا)، كما أشار الإحصائي "نايت سيلفر"؛ مما نتج عنه توقعات انتخابية غير صحيحة. في الانتخابات الرئاسية الأمريكية عام 2008 بين "باراك أوباما" و"جون ماكين"، وجدت مؤسسات توقع الاقتراع الكبرى مثل جالوب وبيو وإيه بي سي" / واشنطن بوست اختلافات تتراوح ما بين 1 % و3 % عندما أخذوا عينة اقتراع تحتوي على مستخدمي الهواتف النقالة وعينات لا تحتوي عليها ـــ وقد كان هذا الاختلاف كبيرًا نظرًا لتقارب نتائج سباق الرئاسة في ذلك الوقت.

الأمر الذي سبب الكثير من المشكلات هو أن العينات العشوائية لم تتمكن من إدراج الفئات الثانوية بسهولة؛ حيث إن تقسيم النتائج إلى مجموعات ثانوية أصغر فأصغر يزيد من احتمالية الوصول إلى توقعات خاطئة. ومن السهل أن نعرف السبب في ذلك. افترض أنك تختار عينة عشوائية مكونة من ألف شخص بشأن من ينوون التصويت له في الانتخابات المقبلة. إذا كانت العينة التي اخترتها عشوائية بشكل كافٍ، فإن احتمالات نسب اختيار الشعب بأكمله للأشخاص أنفسهم بنسب العينة نفسها ستكون في حدود نسبة 3%. ولكن ماذا لو كانت نسبة 3 % بالزيادة أو النقصان غير دقيقة بالشكل الكافي؟ أو ماذا لو أردت أن تقسم المجموعة إلى مجموعات ثانوية أصغر حجمًا طبقًا للنوع أو الجغرافيا أو الدخل؟

وماذا لو أردت أن تدمج هذه المجموعات الثانوية لتستهدف فئة معينة من السكان؟ في عينة تشمل ألف شخص، فإن مجموعة ثانوية على غرار "المصوتات الإناث المترفات من الشمال الشرقي" ستكون أقل من مائة شخص فقط. باستخدام عدد قليل من الملاحظات لتوقع النيات الانتخابية لجميع المصوتات الإناث المترفات من الشمال الشرقي لن تكون تلك التوقعات دقيقة حتى إن كانت العينة عشوائية تمامًا. وستتسبب التحيزات الضئيلة في العينة الإجمالية في أخطاء تظهر بشكل أكبر على مستوى المجموعات الثانوية.

إذن، العينات العشوائية أثبتت أنها لا تفيد عندما ترغب في التعمق أكثر وإلقاء نظرة عن قرب على بعض الفئات الثانوية من البيانات. إن ما يصلح للكم الكبير من البيانات لا يصلح للكم الضئيل. إن العينات العشوائية تشبه إلى حد كبير الطباعة التناظرية للصور، والتي تبدو جيدة من بعيد ولكن عندما تحدق النظر إليها وتكبر جزءًا بعينه من التفاصيل، تبدو الصورة ضبابية.

تتطلب العينات العشوائية أيضًا تخطيطًا وتنفيذًا دقيقين؛ حيث إنه لا يمكن للمرء أن "يطرح" أسئلة مبنية على بيانات حديثة على العينة إذا لم تكن موضوعة في الاعتبار منذ البداية . لذا على الرغم من أن الطرق المختصرة تفيد، إلا أنها غير

مضمونة العواقب. عندما تكون المجموعات البيانية عبارة عن عينات بدلاً من أن تكون كاملة، فإنها تفتقد إلى مستويات معينة من الاتساع والمرونة، بحيث يمكن أن يعاد تحليل البيانات نفسها مرة أخرى بطريقة تختلف تمامًا عن الغرض الرئيسي الذي تم جمعها من أجله.

في حالة تحليل الحمض النووي, وصلت تكلفة عمل تسلسل جيني للفرد الواحد حوالي ألف دولار عام 2012، مما حولها إلى تقنية تصلح للتسويق على مستوى كبير. ونتيجة لهذا، ظهر مجال جديد لعمل تسلسل جيني للأفراد. منذ عام 2007 كانت شركة 23andMe الجديدة في وادي السليكون تقوم بتحليل الحمض النووي مقابل مائتي دولار فقط. كشفت تقنيات الشركة عن صفات متواجدة في الشفرات الوراثية للبشر يمكنها أن تزيد من مقاومتهم لأمراض معينة مثل سرطان الثدي أو أمراض القلب. ومن خلال تكديس بيانات الحمض النووي والحالات الصحية لعملائها، تأمل الشركة في تعلم أمور جديدة لم تكن ظاهرة من قبل.

ولكن توجد عقبة في هذا الأمر وهي أن التسلسل الوراثي الذي تضعه الشركة يمثل جزءًا بسيطًا من الشفرات الوراثية للبشر: الأماكن التي تعمل كمؤشرات على وجود ضعف جيني معين. ولكن ظلت مليارات القواعد الثنائية من الحمض النووي لم يتم وضع تسلسل لها. ولهذا كانت شركة 23andMe وحدها هي القادرة على الإجابة عن الأسئلة المتعلقة بالمؤشرات التي تتعامل معها. عندما يتم اكتشاف مؤشر جديد، يجب أن يتم وضع تسلسل جيني جديد للحمض النووي للشخص (أو على الأقل الجزء المتعلق بالضعف الجيني). إن العمل مع الجزء بدلاً من الكل، من شأنه أن ينشأ عنه أمر جيد: كان بإمكان الشركة أن تجد ما تبحث عنه بسرعة أكبر وبتكلفة أقل، ولكنها لن تكون قادرة على الإجابة على الأسئلة التي لم تتعامل معها قبل هذا.

اتخذ أسطورة شركة أبل ومديرها التنفيذي "ستيف جوبز" أسلوبًا مغايرًا خلال معركته مع مرض السرطان؛ حيث أصبح أحد أوائل الأشخاص في العالم الذين تم عمل تسلسل لحمضهم النووي وكذلك للحمض النووي لورمه السرطاني. ولفعل ذلك، دفع "جوبز" رقمًا مكونًا من ستة أصفار ـــ حوالي عدة مئات من أضعاف المبلغ الذي تتقاضاه شركة 23andMe لتفعل المثل. في المقابل، لم يحصل فقط على عينة من حمضه النووي، أو مجموعة من العلامات، بل حصل على ملف بيانات يحتوي على جميع شفراته الوراثية.

عند اختيار دواء لأحد مرضى السرطان، يأمل الأطباء أن يكون الحمض النووي للمريض مماثلاً بشكل كاف للحمض النووي للمريض الذي أسهم في التوصل لهذا الدواء. ولكن في حالة "ستيف جوبز" كان فريق الأطباء المعالج له قادرًا على اختيار الدواء المناسب له بناءً على البنية الوراثية الكاملة له. عندما يفقد الدواء تأثيره بسبب تحول المرض أو فشل الدواء في علاجه، يمكن للأطباء أن يستخدموا دواءً آخر ـــ كان "جوبز" يدعو هذا الأمر بالتالي: "القفز من زهرة إلى أخرى". وكان يسخر أيضًا قائلاً: "قد أكون من أوائل الأشخاص الذين يتغلبون على مرض سرطان متفشٍّ بهذه الطريقة أو سأكون أحد أحدث من قضى المرض عليهم". رغم أن ما

توقعه قد نحى منحى حزيًا، فإن العملية التي تمت ــ امتلاك جميع البيانات اللازمة، وليس مجرد جزء يسير منها ــ جعلته يقاوم المرض لعدة أعوام.

من البعض للكل

إن العينات هي نمو لحقبة قيود معالجة البيانات التي كان فيها الناس يحاولون قياس أمور كثيرة في العالم ولكنهم كانوا يفتقدون الأدوات اللازمة لتحليل ما قاموا بجمعه. ونتيجة لهذا الأمر، تعتبر العينات من بقايا تلك الحقبة أيضًا؛ حيث إن النتائج القاصرة للحساب والجدولة لم تستمر على المنوال نفسه؛ حيث إن أجهزة المتابعة وبرامج تحديد المواقع على الهواتف المحمولة وأساليب أزرار الضغط على مواقع الإنترنت وموقع تويتر جميعها تجمع البيانات بشكل كبير مما مكن أجهزة الحاسب من معالجة الأرقام بسهولة أكبر.

لم يعد مفهوم العينات يمثل أمرًا منطقيًّا بالقدر الكافي عندما يصل الأمر لمعالجة كم كبير من البيانات؛ حيث إن الأدوات التقنية للتعامل مع البيانات قد تغيرت بشكل جذري، ولكن أساليبنا وعقلياتنا هي ما تتكيف ببطء.

إن العينات تأتي مع كلفة تم الاعتراف بها منذ فترة طويلة ولكن تم تجاهلها. تفتقر العينات للتفاصيل ولكن في حالات أخرى لا تصلح أية طريقة أخرى عدا اختيار العينات، أما في مجالات أخرى، فقد حدث التحول من جمع بعض البيانات إلى جمع أكبر كم ممكن منها، وإن أمكن، جمع جميع البيانات المتوافرة: **ن= الجميع**.

كما رأينا من قبل، فإن استخدام معادلة: **ن=الجميع** تعني التعمق أكثر داخل البيانات: حيث لا يمكن للعينات أن تفعل الأمر ذاته. ثانيًا، تذكر مثالنا السابق عن العينات؛ حيث إنه من المسموح لنا بهامش خطأ يبلغ 3 % فقط عندما نقيس الأمر استقرائيًّا على السكان جميعهم. في بعض الحالات، يكون لا بأس بهامش الخطأ هذا، ولكنك تفقد بعض التفاصيل والتصنيفات والقدرة على إلقاء نظرة عن قرب على مجموعات ثانوية بعينها. إن التصنيف الطبيعي، للأسف، يكون طبيعيًّا. وعادة ما تجد الأمور الشيقة في الحياة في أماكن لا يمكن للعينات أن تكشفها بشكل كامل.

إذن، لا تعتمد نماذج الإنفلونزا من شركة جوجل على عينات عشوائية صغيرة، بل تستخدم بدلاً من ذلك مليارات طلبات البحث عبر الإنترنت في الولايات المتحدة الأمريكية. إن استخدام هذا الكم الكبير من البيانات بدلاً من استخدام عينات صغيرة يحسن من توقع انتشار الإنفلونزا وصولاً إلى مدينة بعينها بدلاً من ولاية بأكملها أو الدولة كاملة. استخدم "أورين إتزيوني" في البداية كمًّا من البيانات بلغ 12 ألف نقطة بيانية كعينة وعملت على الوجه الأمثل، ولكن قام "إتزيوني" بعد ذلك بإضافة كم آخر من البيانات مما حسن من جودة التوقعات أكثر. وفي النهاية، استخدم موقع "فاير كاست" سجلات رحلات الطيران لأغلب الخطوط الجوية لعام كامل. يقول "إتزيوني":"إن البيانات لا تنقطع ــ يمكنك أن تواصل جمعها طوال الوقت، وفي أثناء فعلك هذا الأمر، يمكنك أن تحصل على المزيد والمزيد من الأفكار من هذه الأنماط".

لذا سيكون من الأفضل أن تتخلى عن الطريق المختصر للعينات العشوائية وأن

تهدف إلى الحصول على بيانات أكثر شمولاً بدلاً من ذلك. إن القيام بهذا الأمر يتطلب آليات معالجة وقدرة تخزينية وفيرة وأدوات حديثة لتحليل هذا الكم الوفير من البيانات. يتطلب الأمر أيضًا طرقًا سهلة ورخيصة لجمع البيانات؛ حيث إنها في الماضي كانت تعتبر من المعضلات المكلفة للغاية. ولكن في الوقت الحالي، قلت تكلفة وتعقيد مكونات هذه المعضلة بشكل كبير. إن ما كان في الماضي من الأمور التي تنفرد بها الشركات الكبرى أصبح الآن متوافرًا للأغلبية.

إن استخدام جميع البيانات يجعل من الممكن أن ترصد العلاقات والتفاصيل التي قد تكون مختفية داخل الكم الهائل من المعلومات. على سبيل المثال، البحث عن البطاقات الائتمانية المزورة يحدث من خلال البحث عن البطاقات غير الطبيعية، والطريقة الفضلى للقيام بهذا الأمر هو معالجة جميع البيانات بدلاً من اختيار عينة عشوائية. إن القيم الشاذة هي أكثر أنواع المعلومات إثارة للاهتمام، ويمكنك أن تحددها فقط عند مقارنة كم معين من التعاملات الطبيعية. إنها إحدى اختصاصات البيانات الضخمة. ولأن التعاملات بالبطاقات الائتمانية تحدث طوال الوقت، فيجب أن تحدث التحليلات بشكل فوري هي الأخرى.

شركة Xoom، شركة مختصة في التحويلات النقدية العالمية، وهي من الشركات المدعومة من أسماء كبيرة في عالم البيانات الضخمة، تقوم بتحليل جميع البيانات المتعلقة بالتعاملات التي تقوم بها. دق نظامها ناقوس الخطر في عام 2011 عندما لاحظ أن متوسط عدد التعاملات ببطاقات ديسكفر قد زاد قليلاً في ولاية نيو جيرسي. وضح "جون كونزي"، المدير التنفيذي للشركة، الأمر قائلاً:"إن هناك نمطًا محددًا عندما كان لا يفترض وجود نمط محدد". كانت التعاملات تبدو قانونية إذا أخذت بصورة فردية، ولكن تبين أنها حدثت بواسطة مجموعة إجرامية، وكانت الوسيلة الوحيدة لتحديد القيم الشاذة هي من خلال تحليل البيانات جميعها؛ حيث إن العينات قد لا تتمكن من تحديدها.

إن استخدام كم البيانات كاملاً قد لا يكون مهمة بالغة الصعوبة. قد لا تكون البيانات الضخمة ضخمة في مطلق الأحوال، ولكنها غالبًا ما تكون كذلك. لقد أصبحت نماذج توقع الإنفلونزا من شركة جوجل أكثر دقة بفضل مئات الملايين من النماذج الرياضية التي استخدمت المليارات من النقاط البيانية. إن السلسلة الكاملة للجين البشري تتكون من ثلاثة مليارات قاعدة ثنائية. ولكن العدد المطلق للنقاط البيانية وحدها وحجم المجموعات البيانية لا يحددان الأمثلة السابقة على أنها أمثلة على البيانات الضخمة، بل إن ما يصنفها على أنها بيانات ضخمة هو أنه بدلاً من استخدام الطريق المختصر للعينات العشوائية، استخدم كل من نماذج الإنفلونزا وأطباء "ستيف جوبز" أكبر كم ممكن من البيانات الكاملة.

إن اكتشاف التلاعب بمباريات الرياضة الشعبية الأولى في اليابان، رياضة السومو، مثال جيد على أن معادلة **ن=الجميع** لا يجب بالضرورة أن تكون على بيانات ضخمة. كانت المباريات المقامة جميعها تواجه اتهامات متواصلة بأنها قد شوهت رياضة الأباطرة، وغالبًا ما كانت تُلغى نتائجها. بحث "ستيفن ليفيت"، عالم اقتصاد بجامعة شيكاغو، عن الفساد الكامن في سجلات المباريات طوال عقد كامل

مضى ــ جميعها. وتحدث في بحث رائع نُشر في مجلة *American Economic Review* ، وكرره في كتابه تحت عنوان *Freakonomics*، مع زميل له عن مدى الفائدة الناتجة من فحص هذا الكم الكبير من البيانات.

لقد قاما بتحليل ما يعادل 11 عامًا من مباريات السومو، أي ما يزيد على 64 ألف مباراة، لكي يكتشفوا القيم الشاذة، ووجدوها بالفعل. لقد حدث بالفعل تلاعب بالمباريات؛ ولكن ليس في المباريات التي شك فيها الجميع. بغض النظر عن المباريات النهائية التي قد تم التلاعب بها أو لم يتم، أظهرت البيانات أن هناك أمرًا يحدث خلال المباريات غير الملحوظة التي تحدث في الأدوار النهائية للبطولات والتي قد لا تبدو ذات أهمية كبيرة لأن المصارعين لا يلعبون فيها على الفوز باللقب.

ولكنّ هناك أمرًا واحدًا تختص به رياضة السومو، وهو أن المصارعين بحاجة إلى إجمالي عدد مرات الفوز في الدورة المكونة من 15 مباراة للحصول على ترتيبهم ودخلهم؛ مما قد يؤدي أحيانًا إلى تضارب في المصالح، فعندما يواجه أحد المصارعين حقق نتيجة 7-7 مصارعًا آخر حقق نتيجة 8-6 أو أفضل، سيعني هذا الكثير بالنسبة للمصارع الأول في حين لن يضر بالمصارع الثاني أو يفيده بأية حال من الأحوال. في مثل هذه الحالات، أظهر تحليل الأرقام أن المصارع الذي يحتاج للفوز غالبًا ما يفوز.

ربما كان المصارعون الذين يحتاجون إلى الفوز يصارعون بكل ما يملكون من قوة؟ ربما، ولكن البيانات أسفرت عن وجود أمر آخر يحدث. إن المصارعين الذين على حافة الخطر كانوا يفوزون بنسبة 25 % أكثر من النسبة المعتادة. ومن الصعب أن ترجع السبب في هذا الأمر إلى اندفاع كم كبير من هرمون الأدرينالين في عروقهم وحده. عندما تم تحليل البيانات بصورة أكثر دقة، أظهرت أن المرة التالية التي تواجه فيها المصارعان أنفسهما كان المصارع الذي خسر المباراة السابقة أكثر توقعًا للفوز بتلك المباراة من المباريات التالية التي تظاهرا فيها بأنهما يتصارعان. لذا يبدو أن النصر في المباراة الأولى كان "هدية" من أحد المتنافسين للآخر؛ حيث إن العلاقات بين لاعبي عالم السومو المنغلق وثيقة للغاية.

كانت هذه المعلومات واضحة أمام العيان طوال الوقت. ولكن ربما فشل أسلوب اختيار المباريات عشوائيًّا في الكشف عنها. رغم أنها كانت تعتمد على إحصائيات أساسية، دون العلم بما يجب البحث عنه، لن يتمكن المرء من معرفة أي من العينات يجب استخدامه. على النقيض، تمكن "ليفيت"وزملاؤه من الكشف عن هذه المعلومات باستخدام كم أكبر من البيانات؛ حيث بذلوا أقصى ما في وسعهم من أجل فحص جميع المباريات. إن التحقيق بواسطة البيانات الضخمة يشبه إلى حد كبير صيد الأسماك: لا يكون من الواضح منذ البداية ما إذا كان المرء سيصطاد أي شيء فحسب، بل أيضًا ليس من الواضح **ما** سيمكن للمرء اصطياده.

يجب ألا تتجاوز مجموعات البيانات مساحات تخزينية هائلة؛ حيث إنه في حالة رياضة السومو، لم تتعد المجموعة البيانية بأكملها عدد البايت الخاص بصورة رقمية عادية في هذه الأيام. ولكن لأن ما حدث كان تحليلاً للبيانات الضخمة، فقد بدا الأمر أكثر من مجرد أسلوب عادي لاختيار العينات عشوائيًّا. عندما نتحدث عن البيانات

الضخمة فإننا نعني بكلمة "ضخمة" أنها أقل مما قد يعنيه المعنى المطلق لها لكنها ضخمة بشكل نسبي: نسبة إلى مجموعة بيانية شاملة.

على مدى زمن طويل، كان أسلوب اختيار العينات عشوائيًا طريقًا مختصرًا جيدًا؛ حيث إنها جعلت من مشكلة تحليل البيانات الضخمة أمرًا ممكنًا في الحقبة قبل الرقمية. ولكن عندما يتم تصغير حجم الصور أو الأغاني الرقمية، تُفقد بعض المعلومات عندما يتم أخذ عينة منها. إن امتلاك الكم الكامل من البيانات (أو كم يقترب من الاكتمال) يمنحنا المزيد من الحرية لاستكشافها وفحصها من زوايا متعددة أو التدقيق على بعض أوجهها.

إن التشبيه الملائم في هذه الحالة قد يكون كاميرا "ليترو"، التي لا تلتقط شعاعًا واحدًا من الضوء فحسب، كما يحدث مع الكاميرا التقليدية، بل تلتقط حقلاً كاملاً من الضوء، حوالي 11 مليون شعاعًا من الضوء؛ بحيث يمكن للمصورين فيما بعد أن يحددوا أي عنصر من عناصر الصورة سيركزون عليه في الملف الرقمي. لا حاجة بنا للتركيز على أوجه بعينها منذ البداية؛ حيث إن جمع كل البيانات الممكن جمعها سيمكننا من فعل هذا الأمر فيما بعد. ولأن جميع أشعة الحقل الضوئي قد تم جمعها، فإن الأمر أقرب إلى جمع كل البيانات المتوافرة. والنتيجة، أن المعلومات التي تم جمعها أصبحت أكثر قابلية "لإعادة استخدامها" من الصورة التقليدية؛ حيث يمكن للمصور أن يقرر ما الذي يحتاج للتركيز عليه قبل أن يضغط على زر التصوير.

بالمثل، لأن البيانات الضخمة تعتمد على جميع المعلومات، أو على الأقل على أكبر كم ممكن منها، فهي تسمح لنا برؤية التفاصيل أو استكشاف تحليلات جديدة دون تشويش. ويمكننا أن نختبر فرضيات جديدة على مستويات كثيرة من التفصيل. مثل هذه الجودة هي ما مكننا من اكتشاف تزوير مباريات السومو، وتحديد مناطق انتشار فيروس الإنفلونزا، ومقاومة مرض السرطان من خلال متابعة جزء بعينه من الحمض النووي للمريض. إنها تمكننا من العمل على مستوى مذهل من الوضوح.

لا شك في أنه ليس بالضرورة أن نستخدم البيانات جميعها بدلاً من عينة منها؛ حيث إننا نحيا في عالم محدود المصادر. ولكن يعد استخدام جميع البيانات المتوافرة أمرًا منطقيًّا في عدد متزايد من الحالات، ويجعل من الممكن القيام بهذا الأمر حيث إنه لم يكن ممكنًا من قبل.

أحد المجالات التي تأثرت بشكل كبير بسبب معادلة ن=الجميع هو العلوم الاجتماعية؛ حيث فقدت العلوم الاجتماعية هيمنتها على البيانات الاجتماعية التجريبية؛ حيث إن تحليل البيانات الضخمة قد حل محل استطلاعات الرأي المتخصصة ذات المهارة العالية في الماضي. كانت أنظمة العلوم الاجتماعية تعتمد بشكل كبير على الدراسات والاستقصاءات التي تعتمد على الاستبيانات. ولكن عندما يتم جمع البيانات بشكل سلبي في أثناء قيام الناس بالأمور الطبيعية التي يقومون بها على أية حال، فإن التوجهات القديمة التي تعتمد على اختيار العينات والاستقصاءات قد اختفت من الساحة تمامًا. يمكننا الآن أن نجمع المعلومات التي لم نكن نستطيع جمعها من قبل، سواء كان ذلك بسبب العلاقات التي كشفتها اتصالات الهواتف المحمولة أو التوجهات التي أميط اللثام عنها في موقع تويتر.

الأمر الأهم، أن الحاجة للعينات قد اختفت تمامًا.
"ألبرت لازلو باراباسي"، أحد أهم العلماء في مجال علم نظرية الشبكات، والذي رغب في دراسة التفاعلات بين البشر على مستوى السكان بأكملهم؛ لذا قام هو وزملاؤه بفحص سجل عشوائي لاتصالات هاتفية أجريت عبر الهواتف المحمولة من إحدى شركات المحمول والتي خدمت حوالي خمس عدد سكان إحدى الدول الأوربية ــ جميع سجلات الاتصالات الهاتفية في فترة بلغت أربعة أشهر. وقد كان هذا الأمر التحليل الأول على مستوى مجتمع كامل باستخدام مجموعة من البيانات من صميم معادلة ن=الجميع. نتج عن العمل على هذا المقياس الضخم ومتابعة مكالمات ملايين البشر عبر الهواتف المحمولة على مدار الساعة أفكار جديدة لم تكن لتنكشف باستخدام أية طريقة أخرى.
الأمر المثير للاهتمام، أنه على العكس من الدراسات الصغيرة، اكتشف الفريق أنه إن تمت إزالة أشخاص من شبكة علاقات اجتماعية ما والذين يمتلكون الكثير من الصلات مع المجتمع من حولهم فإن هذه الشبكة الاجتماعية ينخفض مستواها ولكنها لا تفشل. في حين أنه على الجانب الآخر، عندما تزيل أشخاصًا ذوي صلات من مجتمعهم الحالي فإن الشبكة الاجتماعية تنهار فجأة إن لم تتغير بنيتها وتتأثر. كانت تلك النتيجة مهمة ولكنها غير متوقعة. من كان ليتصور أن الأشخاص ذوي الكثير من الأصدقاء أقل أهمية لاستقرار بنية الشبكة من الأشخاص الذين لديهم أصدقاء من خارج الشبكة؟ يدل هذا الأمر على وجود توجه نحو التنوع داخل المجموعة الواحدة وداخل المجتمع ككل.
إننا نميل إلى التفكير في العينات الإحصائية كنوع من الأساسات غير القابل للتغير مثل قواعد الهندسة وقوانين الفيزياء. ولكن يمتد هذا المفهوم لأقل من قرن مضى، وقد تم ابتكاره من أجل علاج مشكلة معينة في لحظة معينة تحت قيود تقنية معينة. لم تعد هذه القيود موجودة في الوقت الحالي بالقدر نفسه. إن اللجوء لاستخدام العينات العشوائية في زمن البيانات الضخمة يشبه ركوب الحصان في طريق للسيارات. يمكننا أن نواصل استخدام العينات في بعض الحالات، ولكن يجب ألا تكون ــ ولن تكون ــ الأمر المهيمن على أسلوب تحليلنا للمجموعات البيانية الضخمة؛ حيث إننا سنهدف بشكل متزايد إلى استخدامها جميعًا.

3
فوضى البيانات

أصبح استخدام جميع البيانات المتوافرة ممكنًا في عدد متزايد من الحالات. ولكن هناك ضريبة لحدوث هذا الأمر؛ حيث إن زيادة الحجم تمهد الطريق لعدم الدقة، فلا شك في أن الأرقام الخاطئة وأجزاء صغيرة من المعلومات الفاسدة تتسلل إلى المجموعات البيانية. ولكن كان الهدف دائمًا أن تتم معاملة هذه الأرقام الخاطئة والمعلومات الفاسدة على أنها مشكلات ومحاولة التخلص منها حيث إننا نستطيع فعل ذلك جزئيًّا. إن الأمر الذي لم نرغب قط في القيام به هو اعتبار تلك الأمور حتمية وأن نتعلم التعايش معها، ويعتبر هذا الأمر أحد التحولات الرئيسية للانتقال من استخدام البيانات الصغيرة إلى البيانات الضخمة.

في عالم البيانات الصغيرة، كان تقليل الأخطاء والتأكيد على الجودة العالية للبيانات هو التوجه الطبيعي والرئيسي. ولأننا كنا نجمع كمًّا محدودًا من المعلومات فقط، فقد كنا متأكدين من أن الأرقام التي لاقينا الأمرين حتى سجلناها كانت على أقصى درجة ممكنة من الدقة. وقد حسنت أجيال متعاقبة من العلماء أدواتهم لجعلها أكثر دقة، سواء كان هذا من أجل تحديد مواقع الأجرام السماوية أو تحديد أحجام الأجسام الدقيقة باستخدام المجهر. في عالم العينات، كان الهوس بالدقة أمرًا بالغ الخطورة؛ حيث إن تحليل كم محدود من البيانات يعني أن الأخطاء ستتضخم بشكل كبير مما يتسبب في تقليل دقة النتائج ككل.

على مدار التاريخ، كانت أعظم إنجازات الجنس البشري تنبع من غزو العالم من خلال قياسه. بدأ السعي نحو الدقة في أوربا في منتصف القرن الثالث عشر عندما تبنى الفلكيون والعلماء القياس الأكثر دقة لقياس الزمان والفضاء ــ وهو ما أطلق عليه المؤرخ "ألفرد كورسبي": "قياس الواقع".

إذا تمكن المرء من قياس ظاهرة ما، فإن الاعتقاد الضمني ينص على أن المرء سيكون قادرًا على فهمها. فيما بعد، اقتصر القياس على الأسلوب العلمي للملاحظة والتفسير: القدرة على القياس والتسجيل وعرض النتائج القابلة لإعادة الإنتاج، كما يقول "لورد كالفن": "أن تقيس أي أن تعرف". أصبح القياس فيما بعد أساس السلطة، كما قال "فرانسيس بيكون" :"المعرفة هي القوة". في الوقت ذاته، علماء الرياضة، الذين أصبحوا فيما بعد خبراء شئون التأمين والمحاسبين، طوروا طرقًا جعلت من جمع البيانات وتسجيلها وإدارتها بدقة أمورًا سهلة.

في فرنسا في القرن التاسع عشر ــ حيث أصبحت فيما بعد أكثر الدول المتقدمة علميًّا في هذه الحقبة ــ تم تطوير نظام قياس لتحديد الأجسام بدقة من أجل قياس الفضاء والوقت وغيرهما، وتبنت دول أخرى النظام نفسه. وكان الأمر يبدو مثل عرض وحدات قياس أولية متفق عليها دوليًّا في حين لم تكن هناك الكثير من المعاهدات الدولية حينها. لقد كان هذا النظام هو بداية عصر القياس. بعد نصف

قرن، في عشرينيات القرن العشرين، حطم اكتشاف الآلات الكمية حلم القياس الشامل الدقيق. وتوجه البشر بأكملهم، إلا عددًا قليلًا من علماء الفيزياء، نحو أسلوب القياس الخالي من الأخطاء الذي انتشر بين المهندسين والعلماء. وتوسع الأمر أكثر في عالم الأعمال؛ حيث إن علماء الرياضة والإحصاء المنطقيين قد بدأوا في التأثير على جميع مجالات التجارة.

في الكثير من المواقف الجديدة التي ظهرت فجأة في الوقت الحالي، كان وضع عدم الدقة ـــ فوضى البيانات ـــ في الاعتبار إحدى السمات الإيجابية وليس أحد مواطن القصور. ويعبر هذا عن علاقة تبادلية، ففي مقابل تخفيف معايير الأخطاء المسموح بها، يمكن للمرء أن يجمع المزيد من البيانات. ليس الأمر أن "الكم الأكبر يتغلب على الجزء"، ولكن في حقيقة الأمر أحيانًا ما يكون الأمر "الكم الأكبر يتغلب على الجودة الأعلى".

هناك أنواع مختلفة من فوضى البيانات لمواجهتها؛ حيث يمكن أن يشير المصطلح إلى حقيقة أن احتمالية الأخطاء تزداد بإضافة المزيد من البيانات. إذن، زيادة القراءات المركزة من بنية البيانات حتى جزء من الألف يزيد من فرصة أن يكون بعض هذه البيانات خاطئًا. ولكن يمكنك أن تزيد أيضًا من فوضى البيانات من خلال المزج بين أنواع مختلفة من المعلومات من مصادر مختلفة، والتي لا تندمج بعضها ببعض بشكل كامل. على سبيل المثال، استخدام برامج التمييز الصوتي لتصنيف المكالمات في أحد مراكز خدمة العملاء، ومن ثم مقارنة هذه البيانات بالوقت الذي يستغرقه الموظف للتعامل مع المكالمة، من الممكن أن ينتج عن هذه الطريقة لمحة خاطفة مفيدة عن الموقف بأكمله ولكنها ليست مثالية. يمكن أن تشير فوضى البيانات أيضًا إلى عدم انتظام الهيئة البيانية، والتي تعني أن البيانات ستكون بحاجة إلى "تنظيف" قبل أن تتم معالجتها. هناك عشرات الآلاف من الطرق التي يمكن الإشارة من خلالها إلى شركة IBM، كما يشير خبير البيانات الضخمة "دي جيه باتيل"؛ حيث يمكن أن نقول شركة IBM أو معامل تي جيه واتسون أو الآلات التجارية الدولية. يمكن أن تنشأ فوضى البيانات عندما نجمع أو نعالج البيانات؛ حيث إننا عندما نفعل هذا الأمر فإننا نقوم بتحويلها وتغييرها إلى شيء آخر، مثلما نفعل عندما نجري تحليلًا للآراء التي تحملها رسائل تويتر لتوقع أي من أفلام هوليود هذا العام سيكون أعلى إيرادًا. إن فوضى البيانات في حد ذاتها فوضوية.

افترض أننا بحاجة إلى قياس درجة الحرارة في حقل من حقول العنب. إذا كنا نمتلك جهازًا واحدًا لقياس درجة حرارة الأرض بأكملها، فإنه يجب أن نتأكد من أنه دقيق ويعمل طوال الوقت: لا يُسمح بفوضى البيانات. وعلى النقيض، إذا كنا نمتلك لكل واحدة من أشجار العنب جهازًا لقياس الحرارة، فسيمكننا أن نستخدم أجهزة أرخص ثمنًا وأقل دقة (طالما أنها لا تظهر انحيازًا نظاميًّا). الأمر هنا أنه في بعض الحالات قد ينتج عن عدد قليل من أجهزة القياس بيانات خاطئة وأقل دقة أو أكثر "فوضى" من نتائج جهاز واحد دقيق. قد تكون أي من القراءات غير صحيحة، ولكن المجموع الكلي للعديد من القراءات من شأنه أن ينتج صورة أكثر شمولاً، ولأن مجموعات البيانات تلك تحتوي على المزيد من النقاط البيانية، فإنها تقدم قيمة

أعظم من شأنها التقليل من تأثير فوضى البيانات التي تحتويها.

افترض الآن أننا سنزيد من تكرار قراءات جهاز القياس الحراري. إذا أخذنا قراءة واحدة كل دقيقة، فسيمكننا أن نتأكد من أن الترتيب الذي ستصلنا بواسطته البيانات سيكون ترتيبًا زمنيًّا تمامًا. أما إذا زدنا من هذا المعدل ليصل إلى عشرة أو مائة قراءة في الثانية الواحدة، فستقل دقة الترتيب. في أثناء انتقال المعلومات في أرجاء الشبكات، قد يصل التقرير عنها متأخرًا أو خارجًا عن النظام المتعارف عليه، أو ربما تُفقد تمامًا في الكم الهائل من المعلومات. ربما لا تتمتع المعلومات بالكثير من الدقة، ولكن كمها الكبير يستحق التغاضي عن عدم الدقة.

في المثال الأول، ضحينا بدقة البيانات في سبيل اتساعها، وفي المقابل حصلنا على كم من التفاصيل لم نكن لنحصل عليه لو لم نقم بهذه التضحية. في الحالة الثانية، ضحينا بالدقة من أجل زيادة التكرار، وفي المقابل حصلنا على تغيير لم نكن لنلاحظه لو لم نقم بتلك التضحية. رغم أننا قد نكون قادرين على التغلب على الأخطاء إذا ما ألقينا بالمزيد من المصادر عليها ــ فرغم كل شيء، هناك 30 ألف معاملة تتم كل ثانية في سوق نيويورك للأوراق المالية مما يعني أن الترتيب الصحيح مهم للغاية ــ فإنه في الكثير من الحالات يكون التغاضي عن الأخطاء أكثر إثمارًا من محاولة منع حدوثها.

على سبيل المثال، يمكننا أن نتقبل بعض فوضى البيانات في مقابل الكم. كما يقول "فورستر"، أحد المستشارين التقنيين، "أحيانًا يكون ناتج جمع 2 + 2 يساوي 3.9، وهذا كافٍ". لا شك في أن البيانات لا يمكن أن تكون خاطئة تمامًا، ولكننا على استعداد للتضحية بجزء من الدقة في مقابل معرفة التوجه العام للأمور. إن البيانات الضخمة تحول الأرقام إلى شيء يعتمد أكثر على الاحتمالية لا الدقة. إن هذا التغيير سيحتاج إلى وقت طويل لاعتياده، وقد أتى محملاً بمشكلات خاصة به والتي سنتناولها لاحقًا في هذا الكتاب. ولكن في الوقت الحالي يستحق الأمر أن نلاحظ أننا سنحتاج دائمًا إلى أن نتقبل فوضى البيانات عندما نزيد من مقياس عملنا.

يمكن للمرء أن يرى تحولاً مماثلاً فيما يتعلق بأهمية المزيد من البيانات المتصلة بتحسينات أخرى في عالم الحوسبة. يدرك الجميع المدى الذي تطورت به قوى المعالجة على مدار السنوات كما توقع قانون "مور"، والذي نص على أن عدد الترانزيستور على الرقاقات يتضاعف كل عامين بصورة كبيرة. نتج عن هذا التطور المستمر أن أصبحت الحواسب أكثر سرعة وذاكراتها أكثر سعة. ولكن يعلم القليلون منا أن المعادلات الحسابية التي تعمل بها أنظمة التشغيل التي نستخدمها قد تطورت هي الأخرى ــ هناك العديد من المجالات التي تطورت أكثر من المعالجات التي تنبأ بتطورها قانون "مور". إن العديد من الفوائد التي عادت على المجتمع من استخدام البيانات الضخمة لم تأتِ من الرقاقات الأكثر سرعة أو المعادلات الأكثر دقة ولكن بفضل المزيد من البيانات.

على سبيل المثال، تغيرت معادلات الشطرنج الرياضية بشكل طفيف على مدار العقود القليلة الماضية؛ حيث إن قواعد الشطرنج متعارف عليها ومحددة لحد بعيد. والسبب في أن برامج الشطرنج على الحاسب أصبحت تمارس اللعبة بشكل أفضل

هو أنها أصبحت تنهي مباراة الشطرنج بصورة أفضل، وهي تقوم بذلك لأن الأنظمة قد تغذت بالمزيد من البيانات. في حقيقة الأمر، تم تحليل جميع التحركات الممكنة في نهاية مباراة الشطرنج عندما يوجد على رقعة الشطرنج ست قطع أو أقل (ن=الجميع)، وتم وضعها في جدول هائل من شأنه أن يملأ تيرابايت كاملًا من البيانات إذا ما تم فك ضغطه؛ مما يتيح لحواسب لعبة الشطرنج أن تنهي مبارياتها دون أية أخطاء، ولا يمكن لأي إنسان أن يهزم هذه الأنظمة.

إن المدى الذي تتغلب به المزيد من البيانات على المعادلات الحسابية الأفضل قد تم إثباته بشكل قوي في مجال معالجة اللغات الطبيعية: الطريقة التي تتمكن الحواسب من خلالها أن تتعرف على الكلمات في أثناء استخدامها في حواراتنا اليومية. بحلول عام 2000، كان الباحثون في شركة مايكروسوف "ميشيل بانكو" و"إريك بريل" يبحثان عن أسلوب لتحسين برامج التصحيح اللغوي كجزء من برنامج Wo الذي تقدمه الشركة. ولم يكونا متأكدين مما إذا كان من المفيد أن يبذلا جهدهما لتحسين المعادلات الحالية أم يحاولا إيجاد تقنيات جديدة أو إضافة سمات دقيقة جديدة. قبل اختيار أي من هذه الخيارات، قررا أن يختبرا ما سيحدث إذا ما قاما بتغذية الأنظمة الحالية بكم كبير من البيانات الإضافية. كانت أغلب معادلات أجهزة التعليم تعتمد على نصوص تحتوي على إجمالي عدد كلمات يبلغ مليون كلمة أو أقل. أخذ "بانكو" و"بريل" أربع معادلات شائعة وغذياها بثلاثة أنظمة هائلة من البيانات الإضافية: 10 ملايين كلمة، ومن ثم 100 مليون كلمة، وفي النهاية مليار كلمة.

كانت النتائج مذهلة؛ فمع إدخال المزيد من البيانات، تحسن أداء المعادلات الأربع بشكل جذري. في حقيقة الأمر، تحسن أداء المعادلة البسيطة التي كانت تؤدى بالشكل الأسوأ مع نصف مليون كلمة عندما تم إدراج مليار كلمة؛ حيث ارتفعت دقتها من 75 % إلى 95 %. وعلى النقيض، تحسن قليلاً أداء المعادلة التي كانت تؤدى بشكل أفضل مع كم ضئيل من البيانات عندما تم إدراج كم أكبر من البيانات فيها، رغم تحسنها بشكل كبير كما حدث مع المعادلات الأخرى؛ حيث ارتفعت دقتها من 86% إلى 94%. كتب كل من "بانكو" و"بريل" في ورقة بحثية عن هذا الموضوع: "تقترح هذه النتائج أنه علينا أن نضع في اعتبارنا العلاقة المتبادلة بين إنفاق الوقت والمال على تطوير المعادلات الرياضية وإنفاقهما على تطوير قواعد البيانات".

لذا فإن الكم الأكبر يتغلب على الكم الأقل، وأحيانًا يتغلب الكم الأكبر على الأكثر ذكاءً. ماذا عن فوضى البيانات إذن؟ بعد بضع سنوات من فحص "بانكو" و"بريل" لهذا الكم الهائل من البيانات، كان الباحثون في شركة جوجل المنافسة يفكرون في الأمر ذاته ــ ولكن على مقياس أكبر؛ حيث كانت جوجل لا تفكر في هذا الأمر من أجل تطوير برنامج للتصحيح اللغوي، بل كانت تسعى لحل أمر أكثر تعقيدًا: ترجمة اللغات.

كانت أجهزة الترجمة حلم رواد الحاسبات منذ فجر عصر الحوسبة في أربعينيات القرن العشرين، عندما كانت الأجهزة مكونة من الأنابيب المفرغة وكبيرة الحجم لدرجة أنها كانت تملأ غرفة بكاملها. أصبحت الفكرة ملحة أكثر خلال فترة الحرب الباردة، عندما حصلت الولايات المتحدة الأمريكية على كم كبير من المواد المكتوبة

والتسجيلات الصوتية الناطقة باللغة الروسية ولكنها لم تكن تملك العدد الكافي من الأيدي العاملة لتتمكن من ترجمتها بسرعة.

في البداية، لجأ علماء الحاسب إلى المزج بين القواعد اللغوية وقاموس ثنائي اللغة، وتمكن أحد حواسب شركة IBM من ترجمة 60 عبارة روسية إلى اللغة الإنجليزية عام 1954، مستخدمًا 250 زوجًا من الكلمات في حصيلة الحاسب اللغوية وست قواعد لغوية. وكانت هذه النتائج واعدة للغاية. تم إدخال جملة "*Mi pyeryedayem mislyi posryedestvom ryechyi*"، في جهاز IBM 701 عن طريق البطاقات المخرمة، وكانت النتيجة ترجمتها إلى الجملة التالية: "إننا نحول الأفكار من خلال وسائل الحديث". وقد تمت ترجمة الجمل الستين جميعها "بسلاسة"، طبقًا لما كُتب في مطبوعة احتفالية أصدرتها شركة IBM بهذه المناسبة. وتنبأ مدير برنامج الأبحاث بجامعة جورج تاون "ليون دوسترت" بأن الترجمة الآلية ستصبح "حقيقة واقعة في خلال خمس أو ربما ثلاث سنوات".

ولكن تبين أن النجاح الأولي ما هو إلا خداع كبير. بحلول عام 1966، اضطرت مجموعة من رواد الترجمة الآلية إلى الاعتراف بالفشل؛ حيث كانت المشكلة أصعب بكثير مما كانوا يتصورون؛ فتعليم الآلات أن تترجم اللغات لا يعني تعليمها القواعد فحسب، بل استثناءاتها أيضًا. إن الترجمة لا تعتمد فقط على الحفظ والاسترجاع، بل تعتمد على اختيار الكلمات المناسبة للموضوع من بين الكثير من البدائل. هل تعني كلمة "*bonjour*"صباح الخير" حقًّا؟ أم تعني "يومًا سعيدًا" أم "مرحبًا" أم "أهلاً"؟ الإجابة هي، حسب الموقف ...

في أواخر ثمانينيات القرن العشرين، طرأت على الباحثين في شركة IBM فكرة جديدة. بدلاً من محاولة تغذية الحواسب بالقواعد اللغوية الواضحة إلى جانب القاموس، سيتركون الحاسب يستخدم أسلوب الاحتمالية الإحصائية لحساب أية كلمة أو عبارة في إحدى اللغات تتفق أكثر مع ما يقبلها في اللغة الأخرى. في تسعينيات القرن العشرين، في مشروع كانديد Candide من شركة IBM، استخدمت الشركة ما يعادل عشر سنوات من النسخ البرلمانية الكندية التي نُشرت باللغتين الإنجليزية والفرنسية ـــ حوالي ثلاثة ملايين زوج من الجمل. ولأنها كانت وثائق رسمية، فقد كانت جودة الترجمة عالية للغاية. وبالنسبة لمعايير الوقت الحالي كانت كمية البيانات المدرجة ضخمة. بعدما اشتهرت تقنية الترجمة الإحصائية الآلية، حولت من تحدي الترجمة اللغوية ليصبح معضلة رياضية كبيرة. وبدا أنها ستنجح. وفجأة، تحسنت الترجمة الحاسوبية بشكل كبير، بعد نجاح هذه القفزة النظرية الهائلة، نجحت شركة IBM في تحقيق القليل من التحسن بصعوبة بالغة رغم إنفاق أموال طائلة. وفي نهاية الأمر، أوقفت الشركة المشروع كاملاً.

ولكن بعد أقل من عقد، في عام 2006، اهتمت شركة جوجل بالترجمة كجزء من مسعاها "لتنظيم المعلومات للعالم وجعلها متوافرة ومفيدة". بدلاً من أن تقوم جوجل بترجمة نصين إلى لغتين، ساعدت جوجل نفسها بكم هائل من البيانات التي تحتوي على الكثير من فوضى البيانات: شبكة الإنترنت العالمية بأكملها وربما أكثر؛ حيث امتص النظام كل التراجم التي تمكن من الوصول إليها من أجل تدريب

الحاسب على الترجمة. حصل النظام على مواقع المؤسسات الكبرى متعددة اللغات والترجمات الدقيقة للوثائق الرسمية والتقارير التي تصدرها الكيانات الدولية مثل الأمم المتحدة والاتحاد الأوربي. حتى تراجم الكتب من مشروع جوجل للمسح الضوئي للكتب كانت مدرجة؛ ففي حين استخدم مشروع كانديد ثلاثة ملايين زوج من الجمل المترجمة بجودة عالية، استخدم نظام جوجل مليارات الصفحات المترجمة مختلفة الجودة، كما يقول مدير مشروع ترجمة جوجل "فرانز جوزيف أوك"، أحد أشهر الخبراء في هذا المجال. وزادت قاعدة بيانات النظام من مليار جملة لتصل إلى 95 مليار جملة إنجليزية وإن كان مشكوكًا في جودتها.

رغم فوضى البيانات التي تحتوي عليها المدخلات، فإن خدمات جوجل هي أنجح الخدمات على شبكة الإنترنت؛ حيث إن تراجمها أكثر دقة من تراجم أنظمة أخرى (ولكنها لا تزال غير مثالية)، وتحتوي على كم أكبر بكثير من المفردات. بحلول منتصف عام 2012، غطت قاعدة بيانات النظام ما يزيد على 60 لغة، ويمكنه أن يتقبل مدخلات صوتية بأربع عشرة لغة للترجمة السلسة. ولأنها تعامل اللغات على أنها بيانات فوضوية وهو ما يجعل الاحتمالات هي العنصر المسيطر، يمكنه أن يترجم بين اللغات الأخرى وبعضها، مثل اللغة الهندية والإسبانية الكتالونية، والتي لا توجد الكثير من التراجم المباشرة بينها ليستخدمها النظام لتطوير نفسه. في هذه الحالات يستخدم النظام اللغة الإنجليزية كجسر. ويعتبر هذا الأسلوب أكثر مرونة من الأساليب الأخرى؛ حيث يمكنه أن يضيف ويزيل الكلمات بمجرد دخولها حيز الاستخدام أو خروجها منه.

إن السبب في نجاح نظام الترجمة من جوجل ليس أنه يحتوي على معادلات رياضية أفضل، بل نجح لأن من ابتكروه، مثلما فعل "بانكو" و"بريل" في مايكروسوفت، قد غذوه بالمزيد من البيانات ــ ليس البيانات ذات الجودة العالية فقط. كانت جوجل قادرة على استخدام كم من البيانات أكبر بمقدار عشرة آلاف مرة من البيانات التي استخدمها مشروع كانديد من مايكروسوفت لأنها كانت تتقبل البيانات الفوضوية. نظام معالجة الكلمات الذي قدمته جوجل في عام 2006 الذي يبلغ قوامه تريليون كلمة كان يتكون من بيانات من الإنترنت من كل حدب وصوب ــ "البيانات الجامحة"، كما يمكننا تسميتها. وكان هذا البرنامج هو "المجموعة التدريبية" التي تمكن النظام من خلالها أن يحسب الاحتمالات التي ــ على سبيل المثال، تمكنه من إدراك أن إحدى الكلمات تتبع كلمة أخرى في اللغة الإنجليزية. وقد كان هذا الأمر بعيدًا كل البعد عن أقدم نظام في هذا المجال، وهو مجموعة بيانات "براون" الشهيرة في ستينيات القرن العشرين، والتي كان مجموع كلماتها مليون كلمة. تسبب استخدام كم أكبر من البيانات في حدوث تقدم كبير في مجال معالجة اللغات الطبيعية والتي اعتمدت عليها أنظمة أخرى مثل التعريف الصوتي والترجمة الحاسوبية. كتب رائد الذكاء الصناعي من شركة جوجل "بيتر نورفيج" في ورقة بحثية تحت عنوان "الفاعلية غير المنطقية للبيانات"، قائلاً: "النماذج البسيطة مع كم كبير من البيانات تتغلب على النماذج الأكثر تعقيدًا مع كم ضئيل من البيانات".

كما وضح "نورفيج" وزملاؤه المؤلفون، لقد كانت فوضى البيانات هي الأساس: "لقد كان هذا الكيان البياني تراجعًا عن مجموعة بيانات براون: لقد أخذناه من صفحات إنترنت غير مرشحة ولهذا السبب يحتوي على جمل غير كاملة وأخطاء إملائية ولغوية وجميع أنواع الأخطاء الأخرى. لم تكتب الحواشي مصححة يدويًّا بعناية، ولكن حقيقة أنه أكبر بمليون مرة من مجموعة بيانات براون تفوق مواطن القصور تلك".

الكم الأكبر يتغلب على الجودة

من الصعب تقبل المحللين التقليديين المتمسكين بأسلوب اختيار العينات للفوضى؛ حيث إنهم ركزوا طوال حياتهم على منع فوضى البيانات واستئصالها. إنهم يعملون بجد لتقليل الأخطاء عند جمع العينات، واختبار العينات المختارة على جميع التوجهات المحتملة قبل إعلان نتائجها. إنهم يستخدمون إستراتيجيات متعددة لتقليل الأخطاء من بينها التأكد من أن العينات التي تم جمعها، قد جُمعت طبقًا لبروتوكول دقيق وبواسطة خبراء مدربين جيدًا. مثل هذه الإستراتيجيات تكلف كثيرًا عند تنفيذها حتى مع كم محدود من البيانات، ومن الصعب تطبيقها على البيانات الضخمة. إنها لا تكون باهظة التكاليف فحسب، بل إن معايير الجمع الدقيقة لا يمكن تحقيقها بشكل متواصل على هذا المقياس. وحتى لو تم استثناء التفاعلات البشرية، فلن يمكن حل المشكلة.

إن دخول عالم البيانات الضخمة سيتطلب منا تغيير تفكيرنا فيما يتعلق بحسنات الدقة. إن تطبيق العقلية التقليدية للقياس على العالم الرقمي المتصل في القرن الحادي والعشرين ما هو إلا حيد عن الطريق الصحيح. كما ذكرنا سابقًا، إن الهوس بالدقة ما هو إلا نتاج العصر التناظري المفتقد للمعلومات. عندما كانت البيانات محدودة، كانت كل نقطة بيانية مهمة، وبالتالي كنا نولي عناية فائقة لتجنب ترك أي منها تؤثر على التحليل.

في الوقت الحالي، نحن لا نعيش في هذا الموقف المحروم من المعلومات. عند تعاملنا مع مجموعات بيانية أكثر شمولاً، والتي تتمكن من اكتشاف ليس فقط أجزاء صغيرة من الظواهر المتوافرة لنا بل أجزاء كثيرة منها وربما كلها، فإننا لا نكون بحاجة إلى أن نقلق كثيرًا على البيانات الفردية التي من شأنها التأثير على سير التحليلات بأكملها. بدلاً من أن نهدف إلى محو كل أثر لعدم الدقة بتكلفة باهظة، فإننا نقوم بالحساب مع وضع فوضى البيانات في الاعتبار.

خذ مثالاً، الطريقة التي تعمل بها الحساسات في المصانع. في مصنع تكرير شركة BP بمنطقة تشيري بوينت في مدينة بلاين بولاية واشنطن، تم تركيب حساسات لا سلكية في جميع أرجاء المصنع لتشكل شبكة غير مرئية تنتج كمًّا هائلاً من البيانات على الفور. وكان من شأن البيئة المحيطة بالحساسات من حرارة عالية وأجهزة إلكترونية أن تربك عمل الحساسات مما يتسبب في إنتاج بيانات فوضوية. ولكن الكم الهائل من البيانات؛ التي جمعتها الحساسات السلكية واللاسلكية قد عوض عن فوضى البيانات هذه، حيث إن زيادة تكرار وعدد مواقع قراءات الحساسات قد

تكون تعويضًا كبيرًا عن البيانات الخاطئة. من خلال قياس الضغط على الأنابيب على مدار الساعة بدلاً من أوقات محددة، تمكنت الشركة من إدراك أن بعض أنواع البترول الخام يسبب التآكل أكثر من الأنواع الأخرى ـــ الأمر الذي لم تتمكن من اكتشافه أو معالجته عندما كان كم البيانات الذي تجمعه ضئيلاً.

عندما يكون كم البيانات ضخمًا ومن نوع جديد، لا تكون الدقة في بعض الحالات هي الهدف طالما سنتمكن من معرفة التوجه العام للأمور. إن الانتقال إلى مستويات القياس الأكبر لا يغير من توقعات الدقة فحسب، بل أيضًا من القدرة العملية على تحقيق الدقة. رغم أن هذا الأمر قد يبدو منافيًا للمنطق للوهلة الأولى، فإن التعامل مع البيانات على أنها شيء غير كامل وغير دقيق يمكننا من القيام بتوقعات مهمة، وبالتالي نتمكن من فهم العالم من حولنا بشكل أفضل.

الجدير بالذكر أن فوضى البيانات ليست أمرًا ملازمًا للبيانات الضخمة، بل هي نتاج عدم كمال الأدوات التي نستخدمها لقياس المعلومات وتسجيلها وتحليلها. إذا أصبحت التقنيات المستخدمة مثالية إلى حدٍّ ما، فسوف تختفي مشكلة عدم الدقة. ولكن طالما كانت تلك الأدوات ليست كاملة، فإن فوضى البيانات ستظل حقيقة عملية علينا التعايش معها، ومن المحتمل أن تظل ملازمة لنا لوقت طويل. إن الجهود المثابرة لزيادة الدقة عادة ما يكون لها تأثير اقتصادي؛ حيث إن قيمة الحصول على كم أكبر من البيانات تفرض نفسها على الساحة. وكما فعل الإحصائيون في الحقبة الماضية عندما لم يهتموا بالعينات الأكبر حجمًا في سبيل العشوائية، يمكننا أن نحيا في ظل القليل من عدم الدقة في سبيل الحصول على المزيد من البيانات.

يمثل مشروع المليار سعرًا استثنائيًّا لهذه النقطة. ينشر المكتب الأمريكي لإحصاءات الأيدي العاملة كل شهر قائمة بأسعار المستهلكين والذي يستخدم لقياس معدل التضخم، وتعتبر هذه الأرقام مهمة للمستثمرين والشركات. ويستعين بها البنك المركزي الأمريكي عند تقرير ما إذا كان سيزيد أو يقلل من معدل الفائدة. وتحدد الشركات زيادة رواتب موظفيها بناءً على معدل التضخم. وتستخدمها الحكومة الفيدرالية لجدولة مدفوعاتها مثل التأمينات الاجتماعية والفائدة التي تدفعها عن صكوك تأمينية معينة.

للحصول على هذه الأرقام، عين المكتب الأمريكي لإحصاءات الأيدي العاملة المئات من فرق العمل لتتصل هاتفيًّا وعبر الفاكس وتقوم بزيارات للمحال والمكاتب في 90 مدينة على مستوى الدولة بأكملها من أجل جمع حوالي 80 ألف سعر لكل شيء من الطماطم وحتى أجرة سيارات الأجرة. ويتكلف جمع هذه البيانات حوالي 250 مليون دولار سنويًّا. ومقابل هذا المبلغ، تكون البيانات دقيقة ومتقنة ومنظمة. ولكن في الوقت الذي تصدر فيه الأرقام يكون قد مر عليها بضعة أسابيع، وكما أظهرت لنا الأزمة الاقتصادية في عام 2008، فإن بضعة أسابيع قد تكون فترة طويلة للغاية. إن صناع القرار بحاجة إلى الحصول على أرقام التضخم بشكل أسرع حتى يتمكنوا من علاج المشكلة بشكل أفضل، ولكن لا يمكنهم الحصول عليها بالطرق التقليدية التي تركز على العينات وما يترتب عليها من دقة.

وبناءً عليه، قام اثنان من علماء الاقتصاد بمعهد ماساتشوستس للتكنولوجيا وهما "ألبرتو كافالو" و"روبرتو ريجوبون"، بابتكار بديل يعتمد على البيانات الضخمة من خلال التحول إلى مسار أكثر فوضوية. باستخدام أحد برامج الحاسب لفحص شبكة الإنترنت، تمكنا من جمع نصف مليون من أسعار المنتجات التي تباع داخل حدود الولايات المتحدة الأمريكية كل يوم. كانت البيانات فوضوية، ولم تكن جميع النقاط البيانية التي تم جمعها قابلة للمقارنة، ولكن من خلال المزج بين جمع البيانات الضخمة والتحليل الماهر، تمكن المشروع من اكتشاف انكماش في الأسعار مباشرة بعد إعلان الإخوة "ليهمان" إفلاسهم في سبتمبر 2008، في حين انتظر من كانوا يعتمدون على قائمة أسعار المستهلكين حتى شهر نوفمبر ليكتشفوا هذا الأمر.

كان مشروع معهد ماساتشوستس للتكنولوجيا ممولاً من مشروع تجاري يُدعى منظمة الأسعار والذي تستخدمه البنوك وغيرها من المؤسسات لاتخاذ القرارات الاقتصادية. يجمع المشروع في الوقت الحالي ملايين المنتجات التي تبيعها مئات المتاجر في أكثر من 70 دولة من جميع أنحاء العالم. لا شك في أن الأرقام المجمعة تحتاج إلى التفسير بعناية، ولكنها أفضل بكثير من الإحصاءات الرسمية التي تشير إلى توجهات التضخم؛ لأنه يحتوي على المزيد من أسعار المنتجات إلى جانب أنها تصدر على الفور مما يمنح صناع القرار ميزة مهمة. (يعمل هذا الأسلوب أيضًا كمرجع خارجي معترف به في الكيانات الإحصائية الدولية، على سبيل المثال، لا تثق جريدة *The Economist* بالطريقة التي تقوم بها دولة الأرجنتين لحساب معدلات التضخم بها؛ لذا فإنها تعتمد على منظمة الأسعار بدلاً من أرقام الأرجنتين).

فوضى البيانات في خضم العمل

في الكثير من المجالات التكنولوجية والاجتماعية، نتعلم أنه من الأفضل الحصول على الأكثر والفوضوي بدلاً من الحصول على الأقل والأكثر دقة. فكر في حالة تصنيف المحتوى. منذ قرون طور البشر تصنيفات وجداول من أجل تخزين المواد واسترجاعها. ولطالما كانت هذه الأنظمة الهرمية غير كاملة، كما يتذكر الجميع بألم كتالوج المكتبات المعتمد على البطاقات، ولكنها نجحت بشكل كافٍ في عالم البيانات المحدودة. ولكن إذا عملت على مقياس أكبر، فإن الكثير من أنظمة الشرائط الممغنطة وكذلك هذه الأنظمة التي تفترض وجود كل شيء في مكانه المحدد بدقة ــ تفشل فشلاً ذريعًا. على سبيل المثال، في عام 2011، كان موقع مشاركة الصور "فليكر" يحتوي على أكثر من 6 مليارات صورة من أكثر من 75 مليون مستخدم. وكانت محاولة تصنيف كل صورة طبقًا لتصنيفات معدة مسبقًا أمرًا غير ذي فائدة. هل سيوجد تصنيف تحت عنوان "قطط تشبه هتلر"؟

بدلاً من هذا، استُبدلت بالتصنيفات آليات كانت أكثر فوضوية ولكنها كانت أكثر مرونة وقابلية للتكيف مع عالم يتطور ويتغير باستمرار. عندما نرفع صورًا على موقع فليكر، نقوم "بوسمها"، أي أننا نلحق بها أي عدد من بطاقات التسمية ونستخدمها لتنظيم المادة والبحث عنها. ويقوم المستخدمون بعمل هذه البطاقات

وإلحاقها بالصور وفقًا لما يريدون: لا توجد أية تصنيفات معيارية أو محددة مسبقًا ولا توجد أية تصنيفات يجب أن نحاكيها، بل يمكن لأي شخص كان أن يضع البطاقات التي يرغب فيها عن طريق كتابتها. واتضح أن بطاقات التسمية هي المعيار الفعلي لتصنيف المحتويات على الإنترنت، وتم استخدامها في مواقع التواصل الاجتماعي مثل تويتر والمدونات وغيرها. لقد سهلت بطاقات التسمية من الإبحار عبر المحتوى الهائل لشبكة الإنترنت ــ خاصة بالنسبة لأمور مثل الصور والمقاطع المصورة وملفات الموسيقى أي الملفات التي لا تحتوي على نصوص مكتوبة وبالتالي لا تصلح معها محركات البحث بالكلمات.

لا شك في وجود بعض بطاقات التسمية التي تُكتب بهجاء خاطئ، وتدل هذه الأخطاء على عدم الدقة ــ ليس في البيانات ذاتها بل في طريقة تنظيمها؛ مما يزعج العقلية التقليدية المدربة على الدقة. ولكن في مقابل الطريقة الفوضوية التي ننظم بها مجموعة صورنا، نحصل على عدد كبير من بطاقات التسمية، وبالتالي إمكانية وصول أكثر عمقًا وسهولة لها. يمكننا أن نمزج بين مسميات البحث بطرق لم نكن لنستطيع أن نقوم بها من قبل. إن عدم الدقة المصاحبة لبطاقات التسمية تعتبر تقبلاً للطبيعة الفوضوية للعالم؛ حيث إنها ترياق للأنظمة الأكثر دقة التي تحاول فرض العقم الزائف على فوضى البيانات الواقعية، متظاهرة بأن كل ما يوجد في العالم يصلح لأن يُدرج في جداول منمقة. هناك في الفضاء وعلى الأرض أمور أكثر مما تحلم به هذه الفلسفة.

تتفاخر الكثير من المواقع الإلكترونية الشهيرة بإعجابها بعدم الدقة بدلاً من مزاعم الدقة. عندما يرى المرء أيقونة على موقع تويتر أو زر "أعجبني" في موقع فيس بوك في إحدى الصفحات الإلكترونية، فإنها تُظهر عدد الأشخاص الذين ضغطوا عليها. عندما تكون الأعداد ضئيلة، تظهر كل ضغطة برقمها مثل "63"، ولكن عندما تزداد الأعداد، تظهر الأعداد تقريبية، مثل "أربعة آلاف". ولا يعني هذا أن النظام لا يعلم العدد الفعلي، بل يعني أنه كلما ازداد المقياس كان عرض الأرقام الفعلية أقل أهمية. هذا إلى جانب أن الكميات قد تتغير بسرعة كبيرة بحيث إن الرقم قد يصبح قديمًا في وقت ظهوره. بالمثل، تُقدم شركة جوجل وقت الرسائل الحديثة بدقة، مثل "منذ 11 دقيقة مضت"، ولكنها تعامل الفترات الطويلة بعدم اكتراث "منذ ساعتين"، كما يفعل موقع الفيس بوك وغيره من مواقع التواصل الاجتماعي.

إن صناعة برمجيات المعلومات والتحليلات قامت منذ بداياتها على العملاء الواعدين "النسخة الوحيدة من الحقيقة" ــ الكلمات الفوضوية الشائعة التي ابتكرها بائعو التكنولوجيا في بدايات القرن الحادي والعشرين. لقد استخدم المديرون التنفيذيون هذه العبارات دون تهكم، وما زال بعضهم يفعل حتى الآن. بهذه العبارات، تعني الشركات أن كل من يدخل لتصفح المعلومات عن أنظمة الشركة التكنولوجية سيحصل على البيانات ذاتها؛ بحيث تتجنب فرق التسويق والمبيعات الصراع على من يملك العملاء المناسبين أو أعداد المبيعات قبل بداية اجتماعاتهم. قد تتوحد اهتماماتهم إذا ما كانت الحقائق ثابتة، وبالتالي يتوحد تفكيرهم.

ولكن تتسبب فكرة "النسخة الوحيدة من الحقيقة" في تغيير شامل ومفاجئ. لقد بدأنا في إدراك أنه ليس من المستحيل وجود نسخة واحدة من الحقيقة فحسب، بل وأن السعي لتحقيقها يعتبر إلهاءً. حتى نحصد فوائد استئناس البيانات على المقياس الكبير، علينا أن نتقبل فوضى البيانات كجزء رئيسي لسير هذه العملية، وليس كأمر علينا أن نتجنبه.

يمكننا أن نرى عدم الدقة تغزو أحد المجالات التي لا يمكن أن تستغني عن الدقة: تصميم قواعد البيانات، حيث إن محركات قواعد البيانات التقليدية تتطلب أن تكون البيانات؛ على درجة كبيرة من التنظيم والدقة. ولم تكن البيانات تُخزن فحسب، بل كانت تقسم إلى "سجلات" تحتوي على نطاقات، ويحتوي كل نطاق على بيانات من نوع وطول محدد. على سبيل المثال، إذا كان النطاق الرقمي يصل طوله إلى سبعة أرقام، فلا يمكن تخزين أرقام تصل إلى عشرة ملايين أو أكثر به. إذا رغب شخص ما في إدخال عبارة "خارج الخدمة" في أحد حقول أرقام الهاتف، فلن يمكنه هذا؛ حيث إنه يجب تغيير بنية القاعدة البيانية حتى يمكنها تخزين هذه المدخلات. وما زلنا حتى الآن نحاول حل مثل هذه القيود على حواسبنا وهواتفنا الذكية عندما يرفض نظام تشغيلها أن يخزن البيانات التي نرغب في حفظها.

كانت الفهارس التقليدية هي الأخرى محددة مسبقًا؛ مما كان يقيد الأمور التي يمكن للمرء أن يبحث عنها. إذا ما قمت بإضافة فهرس جديد، فسيكون عليك أن تبدأه من جديد مما يستغرق وقتًا طويلاً. لقد صُممت قواعد البيانات التقليدية والمفترض أنها منطقية لعالم كانت فيه البيانات نادرة الوجود، ويمكن أن يُعتنى بها بل وسيُعتنى بها جيدًا. إنه عالم يجب أن تكون الأسئلة التي يرغب المرء في إيجاد إجابات لها واضحة منذ البداية، لأن قواعد البيانات قد صُممت من أجل إجابتها ــ وحدها ــ بشكل كافٍ.

ولكن ما زالت تلك الرؤية عن التخزين والتحليل بعيدة كل البعد عن الواقع، حيث إننا في الوقت الحالي نمتلك كمًّا كبيرًا من البيانات متباينة النوع والجودة. نادرًا ما تصلح هذه البيانات للتصنيفات المنمقة المحددة مسبقًا والمعروفة منذ البداية. وعادة ما تطرأ الأسئلة التي نرغب في إجابة لها فقط عندما نجمع البيانات التي نملكها ونعمل على تحليلها.

أدت هذه الوقائع إلى تصميمات جديدة لقواعد البيانات والتي تناقض القواعد القديمة ــ مبادئ السجلات والنطاقات المحددة مسبقًا والتي تعكس أشكالًا هرمية منمقة ومحددة من المعلومات. كانت اللغة المتفق عليها دائمًا في التعامل مع قواعد البيانات هي SQL، أو "لغة الطلبات المركبة". يدل اسمها على محدوديتها، ولكن حدث التحول الكبير في الأعوام الأخيرة نحو لغة يُطلق عليها noSQL، والتي لا تتطلب سجلات مركبة معدة مسبقًا للعمل؛ لأنها تقبل البيانات متباينة الأنواع والأحجام وتسمح بأن يتم البحث فيها بنجاح. في مقابل السماح بفوضى البيانات البنيوية لقواعد البيانات، تتطلب تصميمات قواعد البيانات تلك المزيد من مصادر المعالجة والتخزين. ويعتبر هذا التحول في متناول أيدينا حيث إن تكاليف التخزين والمعالجة قد قَلَّتْ بشكل ملحوظ.

يشرح "بات هيلاند" ــ أحد الخبراء المشهورين في مجال تصميم قواعد البيانات ــ هذا التحول الجوهري في ورقة بحثية تحت عنوان "إذا ما كنت تمتلك الكثير من البيانات، فإن "الجيد منها" يكفي". بعد أن حدد بعض المبادئ الرئيسية للتصميمات التقليدية التي اندثرت بفعل البيانات الفوضوية متعددة المصادر والدقة، يوضح لنا عواقب هذا الاندثار: "لم يعد يمكننا أن نتظاهر بأننا نحيا في عالم نظيف". إن معالجة البيانات الضخمة تتضمن الفقد الحتمي لبعض من المعلومات ــ كما يدعوها "هيلاند": "منقوصة"، ولكنها تعوض عن ذلك عن طريق الحصول على النتائج بسرعة أكبر؛ حيث يستنتج "هيلاند" قائلاً: "لا بأس من حصولنا على بيانات منقوصة ــ هذا بالضبط ما يحتاج إليه عالم الأعمال".

إن التصميم التقليدي لقواعد البيانات يَعِدُ بتقديم نتائج ثابتة طوال الوقت. إذا ما طلبت معرفة الرصيد المتبقي في حسابك، على سبيل المثال، فإنك تتوقع أن تحصل على الرصيد المتبقي بدقة، وإذا طلبت معرفته مرة أخرى بعد ثوانٍ قليلة، فإنك ترغب في أن يمدك النظام بالنتائج ذاتها مفترضًا عدم تغير أي شيء في رصيدك. ولكن بما أن كم البيانات المتوافرة يتزايد وكذلك أعداد المستخدمين الذين يستخدمون النظام، يصبح من الصعب المحافظة على هذا التطابق.

لا يمكنك أن تجد مجموعات البيانات في مكان واحد فقط؛ حيث إنها تميل إلى أن تتوزع على العديد من الأقراص الصلبة والحواسب. لضمان الدقة والسرعة، قد يتم تخزين السجل في مكانين منفصلين أو ثلاثة. وإذا ما قمت بتحديث السجل في أحد الأماكن، فلن تكون البيانات الموجودة في الأماكن الأخرى صحيحة حتى يتم تحديثها هي الأخرى. وحيث إن الأنظمة التقليدية قد تستغرق وقتًا طويلاً حتى تكمل تحديث جميع البيانات، فإن هذا الأمر لا يعد عمليًّا عندما تكون البيانات موزعة على عدد كبير من الحواسب ويتم الدخول على الخادم من قبل المستخدمين عشرات الآلاف من المرات كل ثانية؛ لذا يعتبر تقبل فوضى البيانات في البيانات أحد حلول تلك المشكلة.

اتسم التحول بزيادة شعبية نظام Hadoop، وهو نظام مفتوح المصدر منافس لنظام MapReduce من شركة جوجل والذي يعالج كميات كبيرة من البيانات بنجاح باهر. يعمل النظام وفق أسلوب تقسيم البيانات إلى أجزاء أصغر حجمًا وتوزيعها على عدد من أجهزة الحاسب. يتوقع البرنامج أن أجهزة الحاسب قد تتوقف عن العمل؛ لذا يحتفظ بنسخ كاملة من البيانات. يفترض البرنامج أن البيانات ليست منمقة أو مرتبة ــ في الحقيقة، يفترض أن البيانات على قدر كبير من الضخامة بحيث لا يمكن تنقيحها قبل معالجتها. في حين أن التحليل التقليدي للبيانات يحتاج إلى عملية يُطلق عليها "الاستخراج والتحويل والتحميل" لنقل البيانات إلى الموقع الذي سيقوم بتحليلها، فإن نظام Hadoop يستغني عن تلك التفاصيل؛ حيث يفترض مسبقًا أن كم البيانات المدخلة كبير للغاية لدرجة أنه لا يمكن نقله وأنه يجب تحليلها في مكان وجودها.

لا شك في أن نتائج نظام Hadoop ليست على درجة دقة النتائج نفسها التي تعطيها قواعد البيانات المنطقية: لا تصلح لإطلاق سفينة فضائية أو تأكيد بيانات

حساب مصرفي. ولكن بالنسبة للكثير من المجالات الأقل أهمية التي لا تتطلب الكثير من الدقة، يقوم بالمطلوب بسرعة أكبر من بدلائه. خذ مثلاً مهمات على غرار تحديد قائمة بعملاء من أجل إرسال حملة دعائية لهم. باستخدام نظام Hadoop تمكنت شركة بطاقات الائتمان من تقليل وقت المعالجة لما يعادل عامين من سجلات الاختبار، أي حوالي 73 مليار عملية، من شهر إلى 13 دقيقة فقط. ويعتبر هذا النوع من السرعة نقطة تحول مهمة في عالم الأعمال.

إن تجربة شركة ZestFinance، الشركة التي أسسها مدير المعلومات السابق بشركة جوجل "دوجلاس ميريل" تؤكد الفكرة السابقة؛ حيث إن تقنيتها تساعد المقرضين على تقرير ما إذا كان عليهم أن يقدموا عروض إقراض صغيرة قصيرة المدة لأشخاص ذوي أرصدة مصرفية صغيرة أم لا. ورغم أن التأكيد التقليدي على تفاصيل الحسابات المصرفية كان يعتمد على عدد محدود من المؤشرات على غرار آخر دفعة قسط متأخرة، فإن شركة ZestFinance تقوم بتحليل عدد كبير من المتغيرات "الأكثر ضعفًا". في عام 2012 نشرت الشركة معدل اقتراض افتراضيًّا يبلغ ثلث متوسط المعدل المتعارف عليه في هذا المجال. إن الطريقة الوحيدة لجعل هذا النظام ينجح هو الاقتناع بوجود فوضى البيانات في البيانات.

يقول "ميريل": "من بين الأمور الشيقة أنه لا توجد أية جداول مليئة ببيانات أشخاص كثر ــ يوجد دائمًا كم كبير من البيانات المفقودة". إن المصفوفة التي تجمع منها الشركة المعلومات مصفوفة منقوصة للغاية؛ حيث إن ملف قاعدة البيانات يعج بالخلايا المفقودة؛ لذا تعيد الشركة "لصق" البيانات المفقودة. على سبيل المثال، حوالي 10% من عملاء الشركة المسجلين متوفون ــ ولكن اتضح أن هذا الأمر لا يؤثر على الأقساط المدفوعة. يقول "ميريل" مداعبًا: "بالنسبة للموتى الأحياء، قد يزعم أغلب الناس أنه لن يتم استرداد القروض، ولكن بفضل بياناتنا يبدو أن الموتى الأحياء سيدفعون قروضهم".

في مقابل التعايش مع فوضى البيانات، نحصل على خدمات قيمة للغاية كان من المستحيل حصولنا عليها باستخدام الأساليب والأدوات التقليدية. طبقًا لبعض التقديرات، فإن 5% فقط من جميع البيانات الرقمية "منقحة" ــ أي أنها تصلح للعمل بها في قواعد البيانات التقليدية. دون تقبل فوضى البيانات، تبقى نسبة 95% المتبقية من البيانات غير المنقحة، مثل المواقع الإلكترونية ومقاطع الفيديو، مبهمة. من خلال السماح لعدم الدقة بالعمل، نفتح لأنفسنا مجالاً للعمل في عالم غير مطروق من الأفكار.

لقد قام المجتمع بتحولين ضمنيين أصبحا متأصلين للغاية في طريقة عملنا لدرجة أننا لم نعد نراهما على أنهما تحولات، بل نراهما على أنهما الحالة الطبيعية للأمور. الأول، أننا افترضنا أننا لا نستطيع استخدام المزيد من البيانات؛ لذا لم نستخدم المزيد من البيانات. ولكن اتضح أن القيود لا تتعلق بكم البيانات وأن هناك المزيد منها ينتظر لجمعه من خلال استخدام معادلة ن=الجميع.

التحول الثاني كان بخصوص جودة المعلومات. كان من المنطقي أن نهتم بالدقة

في حقبة البيانات المحدودة التي كنا نجمع فيها كمًّا محدودًا من المعلومات وكان يجب أن تكون دقيقة قدر الإمكان. قد تظل الحال على ما هي عليه في الكثير من الحالات. ولكن بالنسبة للعديد من الحالات الأخرى، أصبحت الدقة الشديدة أقل أهمية من الحصول على نتائج سريعة أو تقدمات سريعة طوال الوقت.

إن الطريقة التي نفكر بها لاستخدام المعلومات بأكملها في مقابل استخدام كم محدود منها، والطريقة التي نقدر بها فوضى البيانات بدلاً من دقتها، ستمتلك تأثيرًا عميقًا على طريقة تفاعلنا مع العالم. لقد أصبحت البيانات الضخمة جزءًا لا يتجزأ عن حياتنا اليومية، إننا كمجتمع قد نبدأ في أن نرغب في فهم العالم من منظور أكبر وأكثر شمولاً عن ذي قبل، كنوع من معادلة ن = جميع العقول. وقد نتسامح مع التشويش والغموض في المجالات التي اعتدنا تحري الوضوح والدقة فيها، حتى إن كان وضوحًا زائفًا ودقة غير كاملة. قد نتقبل هذا الأمر شريطة أن نحصل في المقابل على حس أكبر بالواقعية ــ ما يعادل رسمًا انطباعيًّا يكون فيه كل جزء فوضويًّا إذا ما تم فحصه عن قرب، ولكن من بعيد يمكن للمرء أن يرى صورة كاملة.

إن البيانات الضخمة من خلال تأكيدها على المجموعات البيانية الشاملة والفوضى البيانية، تساعدنا على الاقتراب من الواقع أكثر مما يفعل اعتمادنا على البيانات المحدودة ودقتها. إن جاذبية الكم "الضئيل" و"المؤكد" مفهومة. قد يكون فهمنا للعالم غير كامل, وأحيانًا ما يكون خاطئًا عندما يكون ما يمكننا تحليله محدودًا، ولكنَّ هناك تأكيدًا مريحًا في هذا الأمر، الثبات المطمئن. هذا إلى جانب، بسبب دهشتنا من كم البيانات التي يمكننا جمعها وفحصها، أننا لم نواجه الالتزام ذاته للحصول على كل شيء، ولا نرى كل الأمور من جميع الزوايا الممكنة. وفي الحدود الضيقة للبيانات المحدودة، يمكننا أن نختال بأنفسنا على دقتنا الكبيرة ــ حتى من خلال قياس التفصيلات لأقصى درجة ممكنة، سنظل نفتقد الصورة الأشمل.

في نهاية الأمر، ربما تتطلب منا البيانات الضخمة أن نتغير لأن نصبح أكثر تقبلاً لعدم النظام وعدم الدقة البيانية. إن بنية الدقة التي يبدو أنها تمنحنا اتجاهات جديدة في الحياة ــ أن الكرة المستديرة تسقط في الفتحة المستديرة، أنه لا يوجد سوى إجابة واحدة للمسألة ــ أكثر مرونة مما كنا نعتقد، وأقررنا، بل وآمنا، بأن هذه المرونة تقربنا أكثر من الواقع.

بفضل التغيير الجذري الذي حدث بسبب تغيير أساليب التفكير، قادنا الأمر إلى تحول ثالث يمتلك القدرة على تغيير المعتقدات الرئيسية التي يقوم عليها المجتمع: فكرة فهم الأسباب الكامنة خلف كل ما يحدث، وبدلاً من هذا، كما سنوضح في الفصل التالي، فإن إيجاد العلاقات داخل البيانات والعمل وفقًا لها قد يكون دائمًا أمرًا جيدًا بدرجة كافية.

4

العلاقات التبادلية

عام 1997 عندما توقف لبعض الوقت عن استكمال بحثه لرسالة الدكتوراه عن الذكاء الاصطناعي بجامعة واشنطن ليعمل في شركة إنترنت محلية تبيع الكتب الإلكترونية عبر الإنترنت. كانت الشركة تعمل منذ عامين فقط ولكنها كانت تقوم بعمل جيد؛ حيث قال: "لقد أحببت فكرة بيع الكتب والمعرفة ـــ مساعدة الناس على إيجاد المعرفة التي يريدون الاستمتاع بها". كان اسم المتجر الإلكتروني A، الذي عين "ليندن" في وظيفة مهندس برمجيات ليضمن عمل الموقع الإلكتروني بسلاسة.

لم يكن فريق عمل أمازون مكونًا من تقنيين فحسب، بل استعان المتجر في الوقت ذاته باثني عشر أو يزيد من النقاد الأدبيين والمحررين من أجل كتابة استعراضات للكتب الجديدة والنصح بشراء الجيد منها. على الرغم من أن قصة أمازون يعلمها الكثيرون، إلا أن قلة هم من يتذكرون أن محتواها كان في البداية يكتبه العاملون بأيديهم؛ حيث يقوم المحررون والنقاد بتقييم واختيار الكتب التي يتم عرضها على صفحات الموقع؛ حيث إنهم كانوا مسئولين عما يُسمى "صوت أمازون" ـــ والذي يعتبر أحد مزايا الشركة ومصدر تميزها على منافسيها. في مقال بجريدة *Wall Street Journal*، قيل عن "صوت أمازون" إنهم أكثر نقاد كتب الأمة تأثيرًا؛ حيث إن آراءهم قد أثرت على الكثير من مبيعات الكتب.

ومن ثم، بدأ "جيف بيزوس"، مؤسس الشركة ومديرها التنفيذي، في اختبار فكرة مؤثرة: ماذا لو تمكنت الشركة من اقتراح كتب بعينها على المستخدمين بناءً على المفضلات التسويقية لكل منهم؟ منذ بدايتها، حصلت أمازون على كميات كبيرة من البيانات عن جميع مستخدميها: ماذا يشترون، ما الكتب التي يلقون نظرة عليها فقط دون شرائها، وكم من الوقت قضوه في إلقاء النظرة على تلك الكتب، وما هي الكتب التي قاموا بشرائها بشكل جماعي.

كانت كمية البيانات التي حصلت عليها الشركة هائلة وقامت في البداية بمعالجة البيانات بالطريقة التقليدية من خلال اختيار العينات العشوائية وتحليلها لإيجاد العلاقات المشتركة بين المستخدمين. وكانت النتائج غير دقيقة، فما كان منك إلا أن تشتري كتابًا في بولندا حتى تصاب بوابل كبير من الاقتراحات بكتب من شرق أوروبا. اشترِ كتابًا عن الأطفال وستُغرق بكم هائل من الاقتراحات بكتب مشابهة. يتذكر "جايمس ماركوس"، ناقد الكتب بشركة أمازون من عام 1996 وحتى عام 2، في دراسته *Amazonia* قائلاً: "كانت الشركة تميل إلى أن تعرض عليك كتبًا تختلف قليلاً عن آخر كتاب اشتريته، وكانت الدعاية لا تنتهي. كان الأمر يشبه كما لو كنت قد ذهبت للتسوق مع مجنون القرية".

وجد "جريج ليندن" حلًّا لهذا الأمر؛ فقد أدرك أن نظام المقترحات لم يكن يحتاج

إلى مقارنة مفضلات أشخاص بمفضلات أشخاص أخرين، وهي مهمة كانت تتم بثقل تقني شديد، بل إن كل ما كان النظام بحاجة إليه هو إيجاد العلاقات بين المنتجات نفسها. في عام 1998، طبّق "ليندن" وزملاؤه فكرة ترشيح "من منتج لمنتج" التعاونية، كما يُطلق على هذه التقنية. وقد أحدث التحول في الأسلوب تغيرًا كبيرًا.

ولأن الحسابات كان من الممكن أن تُجرى في وقت مبكر، فإن المقترحات كانت تظهر في لمح البصر. كان الأسلوب متعدد الجوانب أيضًا، فقد كان قادرًا على العمل مع جميع مصنفات المنتجات؛ لذا عندما بدأت أمازون في بيع منتجات أخرى إلى جانب الكتب، كان بإمكانها أن تقترح الأفلام وأجهزة تحميص الخبز أيضًا. وكانت المقترحات أفضل بكثير من ذي قبل لأن النظام كان يستخدم جميع البيانات المتاحة. يتذكر "ليندن" قائلاً: "كانت الدعابة المنتشرة بين مجموعة العمل هي: إذا كان النظام يعمل بكفاءة، فإن كل ما على أمازون القيام به هو أن تقترح عليك كتابًا واحدًا ــ وهو الكتاب التالي الذي ستقوم بشرائه".

يجب على الشركة الآن أن تقرر ما يظهر على الصفحة الرئيسية للموقع. هل يجب عرض المحتويات المنتجة آليًّا مثل المقترحات الشخصية وقوائم الكتب الأكثر مبيعًا، أم المراجعات التي كتبها فريق محرري الشركة المقيمين؟ أو ما تخبرنا به ضغطات المستخدمين، أم ما يخبرنا به النقاد؟ لقد كانت معركة بين أجهزة الحاسب والبشر.

عندما أجرت الشركة اختبارًا يقارن بين تقارير المبيعات التي أعدها المحررون البشريون بتقارير المبيعات التي أنتجتها أجهزة الحاسب، لم تكن النتائج متقاربة. كانت المواد المستقاة من البيانات قد نتج عنها المزيد من المبيعات، فربما لا تدرك أجهزة الحاسب سبب تفضيل المستخدم الذي يقرأ كتب "إرنست هيمنجواي" لشراء كتب "إف. سكوت فيتزجيرالد". ولكن لم يبدُ هذا الأمر مهمًّا؛ فقد كانت المبيعات تزداد بشكل كبير. في نهاية الأمر، عُرض على محرري الشركة نسبة من مبيعاتها ليتغاضوا عن تقييماتهم للكتب المنشورة على الموقع وتم تسريح المجموعة بأكملها. يتذكر "ليندن" قائلاً: "كنت حزينًا للغاية على تسريح فريق المحررين، ولكن البيانات لا تكذب، والتكلفة كانت عالية للغاية".

في الوقت الحالي، يُقال إن ثلث مبيعات شركة أمازون ينتج عن أنظمتها للاقتراح والتخصيص. بفضل هذه الأنظمة، أخرجت الشركة الكثير من منافسيها من حيز المنافسة: ليس فقط المكتبات ومتاجر الموسيقى الكبرى، بل أيضًا منتجي الكتب المحليين الذين اعتقدوا أن لمستهم الشخصية ستحميهم من رياح التغيير. في حقيقة الأمر، تسبب عمل "ليندن" في ثورة في عالم التجارة الإلكترونية؛ حيث إن أسلوبه يعمل به الآن الجميع تقريبًا في هذا المجال. وبالنسبة لشركة Netflix، إحدى شركات تأجير الأفلام عبر الإنترنت، فإن ثلاثة أرباع الطلبات الجديدة التي ترد إليها تكون بسبب المقترحات التي تقدمها. سيرًا على خطى شركة أمازون، أصبحت آلاف المواقع الإلكترونية قادرة على اقتراح المنتجات والمحتويات والأصدقاء والمجموعات دون أن تدرك سبب اهتمام الناس بها.

قد تكون معرفة السبب أمرًا مبهجًا، ولكنها غير مهمة فيما يتعلق بتحفيز المبيعات. فإن معرفة ما يحفز الناس على الشراء أيًّا كان، له صداه. تمتلك هذه الفكرة القوة

لإعادة تشكيل الكثير من الصناعات، ليس فقط التجارة الإلكترونية. كان موظفو المبيعات في جميع المجالات يُطلب منهم دومًا ولوقت طويل فهم ما الذي يدفع الناس للشراء، وأن يدركوا السبب وراء قراراتهم. يتم دومًا تقدير المهارة الاحترافية والخبرة الطويلة فيما يتعلق بهذا الأمر. تُظهر البيانات الضخمة وجود مقاربة أخرى، ربما تكون أكثر واقعية في بعض الأحيان. أظهرت أنظمة اقتراح المبيعات المستحدثة التي قدمتها شركة أمازون العلاقات التبادلية القيمة دون إدراك الأسباب الكامنة خلفها. إن معرفة **الأمر، لا سببه**، تكفي.

التوقعات والميول

إن العلاقات التبادلية تفيد في عالم البيانات الضئيلة، ولكن في سياق البيانات الضخمة فإنها تتألق بشدة؛ حيث إننا من خلالها يمكننا أن نحصل على الأفكار بسهولة وسرعة ووضوح أكبر من ذي قبل.

إن العلاقات المتبادلة، في جوهرها، تحدد كم العلاقات الإحصائية بين قيمتين بيانيتين. تعني العلاقات المتبادلة القوية أنه عندما تتغير القيم البيانية فمن المرجح أن تتغير القيم الأخرى بدورها. لقد رأينا مثل هذه العلاقات التبادلية القوية في نماذج شركة جوجل للأنفلونزا: زيادة عدد الأشخاص في موقع جغرافي بعينه والذين يبحثون عن مفردات بحث معينة على محرك بحث جوجل، يدل على زيادة عدد المصابين بالأنفلونزا في هذا الموقع الجغرافي. أما العلاقات التبادلية الضعيفة فتعني أنه عندما تتغير إحدى القيم البيانية فإن تغيرًا طفيفًا يطرأ على القيم البيانية الأخرى. على سبيل المثال، يمكننا أن نُجري علاقات تبادلية على العلاقة بين طول الشعر والسعادة ونجد أن طول الشعر لا يخبرنا بالكثير عن السعادة.

إن العلاقات التبادلية تدعنا نحلل الظواهر من حولنا، ليس من خلال إلقاء الضوء على طريقة عملها الداخلية بل من خلال تحديد الدليل المفيد الذي يمكننا من تحليلها. لا شك في أن أقوى العلاقات التبادلية لا تكون كاملة، ومن المحتمل أن يتصرف أمران بنفس الأسلوب بمحض الصدفة. قد "تضللنا العشوائية" العبارة التي استعرناها من الخبير التجريبي "نسيم نيكولاس طالب". عند العمل بالعلاقات التبادلية، لا توجد أية تأكيدات بل احتمالات. ولكن إذا كانت العلاقات التبادلية قوية، فإن احتمالية وجود الرابط تكون كبيرة. يمكن للكثير من مستخدمي أمازون أن يشهدوا على هذا الأمر من خلال الإشارة إلى أحد أرفف الكتب المثقل بتوصيات الشركة.

من خلال السماح لنا بتحديد الدليل الجيد للتعرف على الظواهر، تساعدنا العلاقات التبادلية على التعرف على الحاضر والمستقبل: إذا كان المعطى أ عادة ما يحدث بالتزامن مع المعطى ب، فعلينا أن نتابع المعطى ب لتوقع أن المعطى أ سيحدث. من خلال استخدام المعطى ب كدليل يمكننا أن نعرف ما سيحدث مع المعطى أ، حتى وإن كنا لا نستطيع أن نقيس أو نتابع المعطى أ بشكل مباشر. الأمر الأهم، أنها ستساعدنا على توقع ما سيحدث للمعطى أ في المستقبل. لا شك في أن العلاقات المتبادلة لا يمكنها أن تتنبأ بالمستقبل، بل إنها قادرة فقط على توقع ما

سيحدث في المستقبل مع وجود هامش معين للاحتمالية. ولكن تعتبر هذه القدرة قيمة للغاية.

خذ مثالاً حالة شركة وولمارت، أكبر شركة بيع بالتجزئة في العالم، والتي يصل عدد موظفيها 2 مليون موظف وتصل مبيعاتها السنوية إلى 450 مليار دولار ـــ مبلغ يزيد على إجمالي الناتج المحلي لأغلب دول العالم. قبل أن تقدم شبكة الإنترنت هذا الكم الكبير من البيانات، ربما كانت الشركة تمتلك أكبر مجموعة بيانات ضمن المؤسسات الأمريكية. في تسعينيات القرن العشرين، قامت الشركة بثورة في عالم البيع بالتجزئة من خلال تسجيل جميع منتجاتها في صورة بيانات من خلال نظام أطلقت عليه Retail Link، والذي ساعد الموردين على مراقبة معدل وحجم المبيعات والمخزون. من خلال إنشاء تلك الشفافية، مكنت الشركة الموردين من العناية بالمخزون بأنفسهم. في الكثير من الحالات، لا "تحتفظ" الشركة بالمنتج في مخازنها إلى أن يتم بيعه مما يقلل من المخاطر على مخزونها ويقلل من تكاليفه. بعد أن استخدمت وولمات البيانات أصبحت في الحقيقة أكبر متجر يقدم المنتجات بأمانة للمستهلكين في العالم.

ما الذي يمكن لجميع هذه البيانات التاريخية أن تكشف إذا ما تم تحليلها بالطريقة الصحيحة؟ عملت شركة البيع بالتجزئة مع مستخدمي الأرقام الخبراء من شركة Tera، الشركة التي كانت فيما سبق شركة تسجيل الدخل القومي، لكشف العلاقات التبادلية داخل تلك البيانات. في عام 2004، بحثت شركة وولمارت في قواعد بياناتها العملاقة عن التعاملات السابقة: ما المنتج الذي اشتراه كل عميل وما هي التكلفة الإجمالية له، ماذا اشترى المستهلكون إلى جانب كل منتج، الوقت الذي اشترى فيه العميل المنتج، حتى حالة الطقس في ذلك الوقت. من خلال فعل ذلك، اكتشفت الشركة أنه قبل الأعاصير، لا ترتفع مبيعات المصابيح اليدوية فحسب، بل تزداد أيضًا مبيعات منتج Pop-Tarts، وهو عبارة عن وجبة خفيفة سكرية أمريكية للإفطار. عند اقتراب العواصف، كانت الشركة تكدس علبًا من هذا المنتج في مداخل متاجرها إلى جانب مؤن الأعاصير لتسهيل الأمر على العملاء الذين يدخلون ويخرجون من المتجر؛ مما زاد أيضًا من مبيعاتها.

في الماضي، احتاج العاملون في المقرات الرئيسية للشركات إلى الحدس حتى يبدأوا في جمع البيانات واختبار الأفكار. أما الآن، ومن خلال وجود الكثير من البيانات والمعدات الأفضل، أصبحت العلاقات تظهر بسرعة أكبر وبتكلفة أقل. (الأمر الذي يعني أنه يجب على المرء أن يكون أكثر حذرًا: عندما يزداد عدد النقاط البيانية طبقًا لقواعد الجاذبية، سيمكننا أيضًا أن نكتشف المزيد من العلاقات المتبادلة الزائفة ـــ الظواهر التي تبدو ذات صلة حتى وإن كانت ليست كذلك. يحتاج منا هذا الأمر إلى أن نوليه المزيد من الانتباه، عندما نبدأ في تقديره).

قبل ظهور البيانات الضخمة بوقت طويل، أثبت تحليل العلاقات المتبادلة قيمته الكبرى. قُدم هذا المفهوم بواسطة السيد "فرانسيس جالتون"، ابن عم "تشارلز داروين"، في عام 1888، بعدما لاحظ العلاقة المتبادلة بين طول قامة الشخص وطول ذراعيه. إن القواعد الرياضية التي تقبع خلف هذا الأمر مباشرة ومتينة نسبيًّا

ـــ والتي تبين فيما بعد أنها إحدى السمات الرئيسية، والتي ساعدت على جعل العلاقات المتبادلة أحد أكثر القياسات الإحصائية استخدامًا. ولكن قبل ظهور البيانات الضخمة، كانت فائدة العلاقات المتبادلة محدودة. ولأن البيانات كانت نادرة وكان جمعها مكلفًا للغاية، كان الإحصائيون يميلون دائمًا إلى استخدام دليل لها، ومن ثم يجمعون البيانات ذات الصلة ويجرون تحليل العلاقات المتبادلة ليكتشفوا مدى جودة هذا الدليل. ولكن كيف يتم اختيار الدليل المناسب؟

من أجل إرشادهم، كان الخبراء يستخدمون المناهج التجريبية التي تقودها النظريات ـــ الأفكار المجردة عن كيفية عمل أمر ما. بناءً على تلك المناهج التجريبية، قام الخبراء بجمع البيانات واستخدموا تحليل العلاقات المتبادلة للتأكد مما إذا كانت الأدلة المختارة مناسبة أم لا. إذا لم تكن مناسبة، عادة ما كان الخبراء يحاولون ويثابرون لإيجاد أدلة أخرى صالحة، في حالة ما إذا كانت البيانات قد تم جمعها بشكل خاطئ، قبل أن يعترفوا في نهاية الأمر أن المنهج التجريبي الذي بدأوا به أو حتى النظرية التي اعتمدوا عليها كانت مليئة بالأخطاء وتحتاج إلى إصلاحات. المعرفة الناتجة عن هذه المناهج التجريبية التي تعتمد على التجربة والخطأ، كانت تنتج ببطء شديد؛ حيث إن توجهاتنا الفردية والجماعية قد غطت على المناهج التجريبية التي طورناها وعلى طريقة تطبيقنا لها وكذلك على الأدلة التي حصلنا عليها. لقد كانت هذه العملية بطيئة للغاية، ولكنها كانت ناجحة في عالم البيانات الضئيلة.

في عصر البيانات الضخمة، لم يعد يكفي اتخاذ قرارات بشأن أي المتغيرات علينا أن نفحص من خلال الاعتماد على المناهج التجريبية وحدها؛ حيث إن المجموعات البيانية أصبحت أكبر حجمًا والمنطقة الخاضعة للاختبار أصبحت أكثر تعقيدًا. لحسن الحظ، لم تعد الكثير من القيود التي أجبرتنا على المنهجيات التجريبية موجودة هي الأخرى. لقد أصبحنا الآن نمتلك كمًّا كبيرًا من البيانات والكثير من القوى الحاسوبية التي لم نعد بفضلها بحاجة لأن نختار أحد الأدلة أو عددًا منها يدويًّا وفحص كل منها على حدة. إن التحليل الحاسوبي الدقيق يمكنه الآن تحديد الأدلة التي نرغب في الحصول عليها ـــ كما فعلنا مع نماذج توقع الأنفلونزا من شركة جوجل، بعد فحص حوالي نصف مليار نموذج رياضي.

لم نعد بالضرورة بحاجة إلى منهج تجريبي موضوعي صحيح عن المظاهر التي تحدث لنبدأ بفهم عالمنا. ولهذا، لسنا بحاجة لأن نطور فكرة عن المصطلحات التي يبحث عنها الناس عند تحديد وقت ومكان انتشار الأنفلونزا. لسنا بحاجة للتشكيك في الكيفية التي تحدد بها الخطوط الجوية أسعار تذاكر رحلاتها الجوية. لسنا بحاجة للاهتمام بأذواق المشترين من متاجر وولمات. بدلاً من هذا يمكننا أن نُخضع البيانات الضخمة لتحليل العلاقات المتبادلة بها وأن نخبر أنفسنا ما هي المصطلحات البحثية التي تصلح لأن تكون أفضل الأدلة للأنفلونزا، أو ما إذا كانت تذاكر الطيران سترتفع أم لا، أو ما الذي ستشتريه الأسر خلال الأعاصير. بدلاً من أسلوب المناهج التجريبية، يمكننا أن نستخدم أساليب تعتمد على البيانات ذاتها، وسوف تكون نتائجنا أقل تحيزًا وأكثر دقة، وسوف نحصل عليها بسرعة أكبر دون أدنى شك.

إن التوقعات التي تقوم على العلاقات المتبادلة تتمركز في قلب البيانات الضخمة. ويتم استخدام تحليل العلاقات المتبادلة بشكل متكرر لدرجة أننا لا نلاحظ أحيانًا التشتت الذي تسببه، وسوف تزداد الاستخدامات أكثر.

على سبيل المثال، تُستخدم نتائج الأرصدة المالية لتوقع السلوك الشخصي. ابتكرت شركة Fair Isaac Corporation، مفهوم نتائج الأرصدة في خمسينيات القرن العشرين. في عام 2011، أنشأت الشركة "نتائج الالتزام بالعلاج". من أجل تحديد مدى التزام الناس بتناول أدويتهم، حللت الشركة كمًّا كبيرًا من المتغيرات ــ بما فيها متغيرات تبدو أنها غير ذات صلة بالأمر، مثل كم من الوقت يعيش الناس معًا في نفس المنزل إذا ما تزوجوا، وكم يستمرون في العمل معًا، وما إذا كانوا يملكون سيارات. كانت تلك النتائج تهدف إلى مساعدة موردي الخدمات الصحية على توفير المال من خلال إخبارهم بأيٍّ من المرضى يحتاج إلى انتباههم. لا توجد أية علاقة بين امتلاك السيارة وتناول المضادات الحيوية، حيث إن الصلة بينهما هي صلة تبادلية خالصة. ولكن إيجاد مثل هذه الأمور ألهم المدير التنفيذي للشركة بأن يتفاخر في عام 2011 قائلاً: "إننا نعلم ما أنت بصدد فعله غدًا".

اشترك في لعبة العلاقات المتبادلة وسطاء بيانات آخرين، كما تم توثيق ذلك في سلسلة المقالات الرائدة بجريدة *Wall Street Journal* تحت عنوان "ما الذي يعرفونه". تمتلك شركة Experian منتجًا أطلقت عليه فكرة الدخل والذي يستطيع تقدير مستوى دخل الفرد جزئيًّا على أساس تاريخ رصيده المصرفي. لقد طور هذا المنتج من نتائجه من خلال تحليل كم هائل من قواعد البيانات عن تواريخ الحسابات المصرفية في مقابل الضرائب العامة من خدمة الضرائب الداخلية الأمريكية. كان الأمر ليكلف الشركات 10 دولارات عن كل شخص للتأكيد على دخله من خلال استمارات الضرائب، في حين تبيع الشركة توقعاتها للعملاء مقابل أقل من دولار واحد. لذا في الحالات المشابهة لتلك الحالة، سيكون استخدام الأدلة أقل تكلفة من الدخول في التعاملات الروتينية المعقدة من أجل الحصول على الأرقام الدقيقة الحقيقية. بالمثل، يقوم مكتب حسابات مصرفية آخر وهو Equifax ببيع "القدرة على الحصول على جداول المدفوعات" و"جدول المصروفات التقديري"، واللذان وعدا بتوقع الأموال الموجودة في حقيبة أيًّا كان مقدارها.

لقد زادت استخدامات العلاقات المتبادلة أكثر فأكثر؛ حيث درست شركة Aviva، واحدة من كبرى شركات التأمين، فكرة استخدام تقارير الحسابات المصرفية وبيانات التسويق للعملاء بالشركات كأدلة لتحليل عينات من الدم والبول لبعض المتقدمين بطلبات تأمين. كان الهدف هو تحديد الأشخاص الذين يُعتقد أنهم أكثر عرضة للمرض مثل ارتفاع ضغط الدم أو السكر أو الاكتئاب. يستخدم هذا الأسلوب البيانات عن أسلوب حياة الأشخاص والتي تشتمل على مئات المتغيرات مثل الهوايات والمواقع الإلكترونية التي يزورونها والفترة التي يقضونها أمام التلفاز، وكذلك تقديرات لدخلهم المادي.

يعد نظام شركة Aviva، الذي طورته شركة Deloitte للاستشارات، نظامًا ناجحًا في تحديد المخاطر الصحية، وقامت شركات تأمين أخرى مثل Prudential و AIG

بفحص إمكانية استخدام أنظمة مماثلة. الفائدة من هذا النظام أن الأشخاص المتقدمين للحصول على التأمين يتجنبون تحليل عينة من دمهم أو بولهم، وهو أمر مزعج للجميع، والذي يجب على شركة التأمين أن تدفع تكاليفه؛ حيث إن الفحوصات المعملية تتكلف حوالي 125 دولارًا للشخص، في حين تتكلف المقاربة المعتمدة على البيانات وحدها 5 دولارات.

بالنسبة للبعض، قد يبدو هذا الأسلوب غريبًا، لأنه يعتمد على بعض السلوكيات التي لا يوجد أي رباط بينها. يبدو الأمر كما لو أن الشركات قد دست جاسوسًا إلكترونيًّا ليتجسس على جميع ضغطات المستخدمين على الفأرة. قد يتردد المستخدمون في زيارة المواقع الإلكترونية للرياضات الخطرة أو يجلسون على الأريكة متناولين رقاقات البطاطس ويشاهدون المسلسلات على التلفاز إذا ما شعروا بأن هذا الأمر سيزيد من أقساط تأمينهم. لا شك في أن تقييد حرية الناس في التعامل مع المعلومات سيكون له عواقب وخيمة. على النقيض، الفائدة هي أن تجعل التأمين أسهل وأقل تكلفة للحصول على المزيد من العملاء، الأمر الذي يفيد المجتمع بدرجة كبيرة، ناهيك عن شركات التأمين ذاتها.

ربما كان خير مثال، أو ربما كبش الفداء، للعلاقات التبادلية في البيانات الضخمة، هي الشركة الأمريكية للبيع بالتجزئة التي يُطلق عليها Target، والتي اعتمدت على التوقعات القائمة على العلاقات المتبادلة في البيانات الضخمة لسنوات طويلة. في أحد التقارير الاستثنائية، روى "تشارلز دوج"، أحد مراسلي جريدة *New York Times*، كيف تعلم شركة Target أن امرأة ما حامل دون أن تخبرهم هي بذلك. الأمر الأساسي، أنهم يقومون باستغلال البيانات ويتركون العلاقات التبادلية تقوم بعملها.

إن معرفة ما إذا كانت إحدى العميلات حاملاً لأمرٌ مهم بالنسبة لبائعي التجزئة؛ حيث إن الحمل يعتبر نقطة تحول في حياة الزوجين؛ حيث تتغير عاداتهما الشرائية بشكل كبير. يبدأ الزوجان في ارتياد متاجر جديدة ويشتريان من علامات تجارية جديدة. تحول المسوقون بالشركة إلى قسم التحليل بها من أجل استكشاف ما إذا كانت هناك طريقة لاكتشاف حالات الحمل لدى العملاء من خلال أنماط شرائهم.

قام فريق التحليل بالشركة بمراجعة تواريخ الشراء للنساء اللواتي سجلن أسماءهن في مسابقة هدية الطفل الرضيع، ولاحظوا أن هؤلاء النساء قد اشترين كميات كبيرة من الغسول عديم الرائحة خلال الشهر الثالث من الحمل، وبعد عدة أسابيع كن يشترين أشياء إضافية مثل الماغنسيوم والكالسيوم والزنك. في النهاية، تمكن الفريق من تحديد دستتين من المنتجات التي يمكن أن تعمل كأدلة تساعد الشركة على حساب "التوقع بالحمل" لكل عميل قام بدفع فاتورته بالبطاقة الائتمانية أو بطاقة العملاء المميزين أو بكوبونات البريد الإلكتروني. يمكن للعلاقات المتبادلة أيضًا أن تدع بائعي التجزئة يقدرون تاريخ الولادة التقريبي؛ لذا يمكن أن يرسلوا الكوبونات ذات الصلة بكل من مراحل الحمل. ويعد هذا "هدفًا" دون أدنى شك.

في كتابه قوة العادات*، يسرد ما حدث بعد ذلك. في أحد الأيام اندفع أحد العملاء غاضبًا داخل أحد متاجر الشركة وطلب مقابلة المدير. صاح العميل قائلاً: "لقد

وصلت هذه الكوبونات إلى ابنتي عبر البريد الإلكتروني، إنها مازالت في المرحلة الثانوية ووصلت لها كوبونات لملابس ومستلزمات الأطفال؟ هل تحاولون تشجيعها على الحمل في هذه السن؟". عندما اتصل المدير بالرجل بعد بضعة أيام ليعتذر له عما حدث، كان صوت الرجل على الجانب الآخر يبدو مستعطفًا؛ حيث قال: "لقد تحدثت مع ابنتي، ويبدو أن هناك أمورًا تحدث في بيتي لم أكن على دراية بها، وسوف تلد في شهر أغسطس، إني أدين لك بالاعتذار".

إن إيجاد الأدلة في السياقات الاجتماعية يعد طريقة واحدة لتوظيف تقنيات البيانات الضخمة، وتمتلك العلاقات التبادلية نفس درجة القوة مع أنواع البيانات الجديدة فيما يتعلق بتلبية الاحتياجات اليومية.

من بين هذه الأساليب أسلوب يُطلق عليه التحليلات التوقعية، والذي بدأ استخدامه على مدى واسع في الشركات لتوقع الأحداث قبل حدوثها. يمكن أن يشير المصطلح إلى معادلة رياضية يمكنها تحديد أغنية ناجحة، الأسلوب الذي يستخدم بشكل شائع في مجال الموسيقى لإعطاء أفكار للمنتجين الموسيقيين عن أي الأماكن سيكون فيها ألبومهم الجديد أكثر رواجًا. يستخدم الأسلوب ذاته لتجنب فشل المشروعات الميكانيكية أو الإنشائية الضخمة: وضع الحساسات على الأجهزة أو المحركات أو البنى التحتية مثل الكباري يسمح بمتابعة أنماط البيانات التي تصدر عنها، مثل الحرارة والترددات والأصوات؛ لتحديد التغيرات التي من شأنها الإشارة لحدوث مشكلات في المستقبل.

إن المفهوم الكامن في هذا الأمر هو أنه عندما تتعطل الآلات، فإنها لا تتعطل فجأة وبدون مقدمات، بل إن هذا يحدث بالتدريج وبمرور الوقت. يمكن لتحليل العلاقات التبادلية وغيرها من الأساليب، المسلحة بالبيانات من الحساسات، أن تحدد أنماطًا معينة وكذلك المؤشرات الدالة التي عادة ما تظهر قبل تعطل آلة ما ـــ طنين الموتور أو الحرارة الزائدة المنبعثة من المحرك، وغيرها. منذ تلك اللحظة، أصبح كل ما يحتاج إليه المرء هو البحث عن هذه الأنماط لمعرفة متى ستتعطل الآلة. إن تحديد الأمور غير الطبيعية مبكرًا يسمح للنظام بأن يرسل تحذيرًا ليلفت نظر المسئولين لاستبدال إحدى قطع الغيار أو إصلاح المشكلة قبل تعطل الآلة. إن الهدف هو تحديد ومراقبة دليل جيد، ومن ثم توقع ما سيحدث في المستقبل.

استخدمت شركة الشحن يو بي إس التحليلات التوقعية منذ بدايات القرن الحادي والعشرين لمراقبة أسطولها المكون من 60 ألف مركبة في الولايات المتحدة الأمريكية لتعرف متى تُجري الصيانة الوقائية؛ حيث يمكن لتعطل إحدى المركبات على الطريق أن يتسبب في خسائر كبيرة، تأخير البضائع المراد توصيلها وبالتالي تأخير تحميل بضائع غيرها. لذا لكي تكون الشركة على استعداد، كانت تقوم باستبدال قطع غيار بعينها كل عامين أو ثلاثة. ولكن لم يكن هذا الأمر كافيًا؛ حيث إن بعض قطع الغيار التي تم استبدالها كانت بحالة جيدة. ومنذ أن تحولت الشركة لاستخدام التحليل التوقعي، تمكنت من توفير ملايين الدولارات من خلال قياس ومتابعة قطع الغيار كل على حدة ومن ثم استبدالها عند الحاجة فقط. في إحدى الحالات، تمكنت البيانات من كشف أن مجموعة كاملة من المركبات الجديدة بها

عيب في إحدى قطعها كان من شأنها أن تسبب الكثير من المشكلات لو لم يتم اكتشافها قبل العمل بها.
بالمثل، يتم تثبيت الحساسات على الجسور والمباني لمراقبة مؤشرات عوامل التعرية، وتستخدم أيضًا في مصانع الكيماويات الكبرى ومحطات تكرير البترول، التي يمكن لآلة واحدة معطلة فيها أن توقف الإنتاج بأكمله. إن تكلفة جمع وتحليل البيانات التي تنبه للقيام بالإصلاح مبكرًا أقل من تكلفة الإصلاح بعد تعطل الآلة. لاحظ أن التحليل التوقعي قد لا يفسر سبب المشكلة، بل يشير فقط إلى وجود المشكلة؛ حيث سيشير فقط إلى أن المحرك تنبعث منه حرارة زائدة, ولكنه لن يحدد لك ما إذا كان السبب هو سير مروحة مهترئ أم مسمار لولبي غير مربوط جيدًا. تشير العلاقات التبادلية إلى المشكلة لا إلى سببها، ولكن كما رأينا من قبل، إن معرفة المشكلة تكفي عادة.
تم تطبيق المنهجية ذاتها على الخدمات الصحية، من أجل تجنب تعطل آلة الجسم البشري. عندما تثبت المستشفى عددًا كبير من الأنابيب والأسلاك والأجهزة بالمريض، يصل للأطباء كم كبير من البيانات. إن جهاز الصورة البيانية الكهربائية للقلب وحده يسجل ألف قراءة كل ثانية. ولكن الجدير بالذكر أن جزءًا يسيرًا فقط من تلك البيانات هو ما يتم استخدامه أو الاحتفاظ به؛ حيث يتم التخلص من معظمها، حتى وإن كانت تحتوي على أدلة هامة على حالة المريض أو استجابته لنوع معين من الأدوية. إن تم الاحتفاظ بتلك البيانات ودمجها مع بيانات مريض آخر، فقد تكشف عن أفكار استثنائية عن أي أساليب العلاج سينجح وأيها سيفشل.
إن التخلص من البيانات قد يكون صحيحًا إذا ما كانت تكلفة وتعقيد جمع البيانات وتخزينها وتحليلها عالية، ولكن ليس الحال هكذا في الوقت الحالي. تعمل الطبيبة "كارولين ماكجريجور" ضمن فريق من الباحثين بجامعة أونتاريو بمعهد التكنولوجيا بمشاركة شركة آي بي إم وعدد من المستشفيات للتوصل إلى برنامج يساعد الأطباء على اتخاذ قرارات أفضل بشأن تشخيصات أمراض الأطفال الرضع غير كاملي النمو (الذين يُطلق عليهم "الأطفال المبتسرون"). يعمل البرنامج على جمع وتحليل البيانات على الفور ومتابعة 16 مصدرًا للبيانات في الوقت ذاته مثل معدل ضربات القلب ومعدل التنفس ودرجة الحرارة وضغط الدم ومستوى الأكسجين بالدم والتي تُصدر مجتمعة حوالي 1260 نقطة بيانية كل ثانية.
يمكن للنظام أن يحدد أدق التغيرات في حالة هؤلاء الأطفال، والتي تدل على بداية الإصابة بالعدوى قبل 24 ساعة من ظهور الأعراض. تشرح الطبيبة "ماكجريجور" الأمر قائلة: "لا يمكنك أن ترى هذه الأعراض بالعين المجردة ولكن الحاسب يمكنه هذا". إن النظام لا يعتمد على السببية بل على العلاقات التبادلية. إنه يخبرنا بالمشكلة لا بسببها. ولكنه يقوم بمهمته على الوجه الأمثل؛ حيث إن الإنذار المبكر يمكن الأطباء من علاج العدوى مبكرًا دون تدخل طبي كبير، أو يخبرهم مبكرًا بأن أسلوب العلاج المتبع غير فعال؛ مما يحسن من حالة المريض. من الصعب أن تفكر في أن هذه التقنيات لن يتم تطبيقها على عدد أكبر من المرضى في المستقبل. قد لا تكون المعادلات الرياضية وحدها هي ما يتخذ القرار، ولكن

تقوم الآلات بأفضل ما يمكن للآلات القيام به، أن تساعد مقدمي العناية من البشر على أن يقوموا بأفضل ما يمكنهم القيام به.

المثير للدهشة، أن تحليل البيانات الضخمة الذي قامت به الطبيبة "ماكجريجور" كان قادرًا على تحديد العلاقات التبادلية بطرق كانت جلية لمعرفة الأطباء التقليديين؛ حيث وجدت، على سبيل المثال، أن هناك مؤشرات حيوية ثابتة تظهر دائمًا قبل حدوث عدوى خطيرة. وهذا غريب؛ حيث إننا قد نشك في أن انتكاس المؤشرات الحيوية يحدث عادة قبل ظهور العدوى. يتخيل المرء أجيالًا من الأطباء ينهون يوم عملهم من خلال ذاكرة تخزين مؤقت بجانب أسرة المرضى؛ حيث يرون استقرار المؤشرات الحيوية لأحد الأطفال، وأن يقرروا أنه من الآمن أن يذهبوا لمنازلهم ــ حتى يصلهم اتصال مفزع من الممرضات في منتصف الليل لتخبرهم بأن أمرًا ما خاطئ قد حدث وأن حدسهم قد خاب.

تقترح الطبيبة "ماكجريجور" أن استقرار المؤشرات الحيوية للأطفال المبتسرين، ناهيك عن تحسن حالتهم، قد تكون الهدوء الذي يسبق العاصفة ــ كما لو كان جسمهم الهزيل يعد أعضاءه للحالة العصيبة التي سيمر بها فيما بعد. لا يمكننا أن نتأكد: فما تشير إليه البيانات هو علاقات تبادلية، وليست سببية. ولكننا على يقين من أنها احتاجت لأساليب إحصائية لتطبيقها على كمية هائلة من البيانات لكشف هذه العلاقات التبادلية الخفية. أما زال هنا أي شك في أن: البيانات الضخمة تنقذ حياة البشر.

الأوهام والتنوير

في عالم البيانات المحدودة؛ حيث يوجد كم محدود من البيانات، بدأت كل من التحليلات السببية والتبادلية بمنهجيات تجريبية، والتي كانت تُختبر من أجل تحديد ما إذا كانت زائفة أم صحيحة. ولكن لأن كلا الأسلوبين كانا يحتاجان إلى مناهج تجريبية للبدء، كان كلاهما عرضة للحكم المسبق أو النوايا السيئة. وكانت البيانات الضرورية غير متوفرة طوال الوقت. أما الآن، مع وجود هذا الكم الكبير من البيانات والمزيد منها في طريقه للظهور، لم تعد تلك المناهج التجريبية ضرورية لتحليل العلاقات التبادلية.

هناك اختلاف آخر بدأ مؤخرًا يحظى ببعض الاهتمام. قبل ظهور البيانات الضخمة، بسبب عدم وجود القوة الحاسوبية الكافية، كانت تحليلات العلاقات التبادلية باستخدام المجموعات البيانية الكبيرة ذات قدرة محدودة على البحث عن العلاقات الخطية. في حقيقة الأمر، إن الكثير من العلاقات معقد للغاية، وباستخدام التحليلات الأكثر دقة يمكننا تحديد العلاقات اللاخطية داخل البيانات.

على سبيل المثال، على مدى أعوام طوال اعتقد الاقتصاديون والسياسيون أن السعادة والدخل توجد فيما بينهما علاقة تبادلية مباشرة: زد من الدخل وستزداد سعادة المواطن العادي. عند النظر إلى البيانات على المخطط، تكشف لنا وجود ديناميكيات أكثر تعقيدًا تشارك في هذا الأمر. إذا ما قلت مستويات الدخل عن حد معين، فإن أية زيادة في الدخل تساوي زيادة تقابلها في السعادة، ولكن بعد هذا

الحد، فإن أية زيادة في الدخل لا تؤثر كثيرًا على سعادة الفرد. إذا ما رغبنا في عرض هذه البيانات في رسم بياني، فسوف يظهر الخط في صورة منحنى بدلاً من خط مستقيم كما يفترض المحللون الخطيون.

كانت النتائج مهمة بالنسبة لصناع السياسات. إذا ما كانت العلاقة خطية، فمن المنطقي أن يتم رفع دخل الجميع لزيادة السعادة الإجمالية للشعب. ولكن بمجرد تحديد أن العلاقة ليست خطية، تغيرت النصيحة للتركيز على زيادة دخل الفقراء؛ حيث إن البيانات أظهرت أن هذا من شأنه التأثير بشكل كبير على سعادتهم.

ويزداد الأمر تعقيدًا، مثلما كانت العلاقات التبادلية تحتوي على أكثر من جانب. على سبيل المثال، درس الباحثون من جامعة هارفارد ومعهد ماساتشوستس للتكنولوجيا التباين فيما يتعلق بالمناعة من مرض الحصبة بين المواطنين: تلقت بعض المجموعات لقاحًا ضد المرض في حين لم تتلقَّ مجموعة أخرى اللقاح. في البداية بدا كما لو أن هناك علاقة تبادلية بين هذا التباين وعدد الأشخاص الذين يذهبون لمراكز الرعاية الصحية. ولكن النظرة عن كثب كشفت أن العلاقة بينهما ليست خطًّا مستقيمًا، بل هي منحنى غريب الشكل. مع صرف الناس للمزيد من الأموال على الرعاية الصحية، يبدأ التباين المناعي في الانخفاض (كما هو متوقع)، ولكن مع صرفهم للمزيد من الأموال، فإن التباين المناعي يعود ليزيد مرة أخرى ــ يبدو أن بعض الأغنياء يخافون من تلقي الجرعات المضادة للحصبة. بالنسبة للمسئولين عن الرعاية الصحية، يعتبر هذا الأمر من الأمور الضرورية ليعرفوها، ولكن لم يتمكن التحليل الخطي البسيط للعلاقات التبادلية من اكتشاف هذا الأمر.

بدأ الخبراء في الوقت الحالي في تطوير الأدوات الضرورية لتحديد ومقارنة العلاقات التبادلية اللاخطية. وفي الوقت ذاته، بدأ تعزيز وتطوير تقنيات تحليلات العلاقات التبادلية من خلال مجموعة من الأساليب والبرامج المستحدثة التي يمكنها كشف العلاقات غير السببية في البيانات من مختلف الزوايا ــ بالطريقة ذاتها التي يحاول الفنان التكعيبي من خلالها أن يلتقط صورة لوجه امرأة من عدة زوايا في اللحظة ذاتها. أحد الأساليب الحيوية الجديدة يمكن إيجاده في مجال تحليل الشبكات المزدهر، والذي يجعل من الممكن تخطيط وقياس وحساب الروابط والعلاقات بين كل شيء بدءًا من صداقات موقع فيس بوك، إلى أي من الأحكام القضائية السابقة عليك الاستشهاد بها، إلى من يتصل بمن على هاتفه الخلوي. هذه الأدوات مجتمعة من شأنها المساعدة على إجابة الأسئلة التجريبية غير السببية.

في النهاية، في عصر البيانات الضخمة، ستقودنا هذه الأنواع الجديدة من التحليلات إلى موجة جديدة من الأفكار الجديدة والتوقعات المفيدة. وسوف نرى صلات لم نتمكن من رؤيتها من قبل. وسوف نمتلك الآليات التقنية والاجتماعية التي طالما ثبطت من مفاهيمنا ناهيك عن أقصى جهودنا. ولكن الأمر الأهم، ستساعد هذه التحليلات غير السببية استيعابنا للعالم من خلال البحث عن المشكلة لا سببها.

في البداية، قد يبدو هذا الأمر منافيًا للمنطق؛ فرغم كل شيء، نحن كبشر نرغب في استيعاب العالم من حولنا بصورة منطقية من خلال العلاقات السببية؛ حيث إننا نرغب في الاعتقاد بأن كل أمر يحدث له سبب إذا ما ألقينا عليه نظرة مدققة. ألا

يجب أن يكون هذا الأمر أقصى أمانينا، أن نعرف الأسباب الخفية وراء كل ما يحدث في العالم؟
لا شك في وجود جدل فلسفي تمتد جذوره لقرون ماضية حول ما إذا كانت السببية موجودة من الأساس. إذا كان كل أمر يحدث بسبب أمر آخر، فإن المنطق يقول إننا لا نملك حريتنا لاتخاذ أية قرارات بشأن أي شيء. لم تكن الإرادة البشرية لتتواجد؛ حيث إن كل قرار نتخذه وأية فكرة تطرأ علينا قد تسبب فيها أمر آخر، والذي بدوره قد حدث بسبب أمر آخر، وهكذا. إن مسار الحياة بأكمله سيتحدد من خلال أسباب تؤدي إلى أحداث. لذا تخاصم الفلاسفة بسبب الدور الذي تلعبه السببية في عالمنا، وفي بعض الأحيان كانوا يقللون من قدرها في مقابل الإرادة الحرة. إن هذا الجدل المجرد لا يعنينا في شيء في هذا الكتاب.
عندما نرى البشر يرون العالم من حولهم من خلال السببية، فإننا نشير إلى طريقتين رئيسيتين يشرح ويفهم البشر من خلالهما العالم: من خلال السببية السريعة الواهمة، أو من خلال التجربة السببية البطيئة المنهجية. سوف تغير البيانات الضخمة من قواعد كل منهما.
في البداية، رغبتنا في رؤية الصلات السببية؛ حيث إننا نميل إلى افتراض وجود الأسباب حتى وإن لم تكن موجودة. ولا يرجع هذا الأمر إلى ثقافتنا أو تنشئتنا أو مستوى تعليمنا، بل، كما تقترح الأبحاث، يتعلق الأمر بالطريقة التي يعمل بها الإدراك البشري. عندما نرى حدثين يقعان أحدهما يلي الآخر، تُجبر عقولنا على أن ترى الأمر بالطريقة السببية.
انظر إلى الجمل الثلاث التالية: "وصل والدا فريد متأخرين، وكان فريد يتوقع حضور متعهدي الطعام مبكرًا، كان فريد غاضبًا". عندما نقرأ تلك الجمل، نعرف على الفور سبب غضب "فريد" ـــ إنه ليس غاضبًا بسبب تأخر متعهدي الطعام، بل بسبب تأخر والديه. في حقيقة الأمر، لا يمكننا أن نعرف أبدًا هذا الأمر من المعلومات المتوفرة لنا. ولكن مازالت عقولنا غير قادرة على المساعدة على الافتراض بأن هذه القصص منطقية وسببية بالمعلومات المتوفرة لنا.
"دانييل كانمان"، أستاذ علم النفس بجامعة برينستون والحاصل على جائزة نوبل في الاقتصاد عام 2002، استخدم هذه الأمثلة لافتراض أننا نمتلك نوعين من التفكير. الأول سريع ولا يحتاج إلى الكثير من الجهد ويسمح لنا بالقفز إلى الاستنتاج في ثانية واحدة. أما الثاني فبطيء وشاق ويتطلب منا التفكير بعمق في موضوع بعينه.
إن الأسلوب السريع في التفكير يميل بشدة نحو "رؤية" الصلات السببية حتى وإن لم تكن موجودة. إنه عبارة عن حكم مسبق للتأكيد على معرفتنا ومعتقداتنا الموجودة. في التاريخ القديم، ساعدنا هذا التفكير على النجاة من البيئات الخطرة، حيثما كنا بحاجة دائمًا إلى اتخاذ قرارات سريعة دون الحصول على قدرٍ كافٍ من المعلومات. ولكنها عادة ما كانت تفشل في تحديد المسبب الحقيقي لأمرٍ ما.
لسوء الحظ، كما يقول "كانمان"، إن عقولنا عادة ما تكون متسرعة لدرجة تجنبها للتفكير المتريث والمنهجي في الأمور، وبدلاً من هذا فإننا ندع التفكير السريع يستحوذ علينا. وينتج عن هذا أن "نرى" طوال الوقت السببيات التخيلية وبالتالي

نفهم العالم بطريقة خاطئة.
عادة ما يقول الآباء لأطفالهم إنهم أصيبوا بالأنفلونزا لأنهم لم يرتدوا قبعات أو قفازات في الأيام الباردة. ولكن لا توجد أية صلة سببية بين عدم ارتداء الملابس الثقيلة والإصابة بالأنفلونزا. إذا ما تناولنا الطعام في أحد المطاعم ثم شعرنا بالمرض بعد ذلك، فربما ألقينا اللوم على الطعام الذي تناولناه في هذا المطعم (وربما تجنبنا دخول هذا المطعم في المستقبل)، حتى وإن لم تكن هناك علاقة بين الطعام الذي تناولناه ومرضنا. قد نكون قد أصبنا بجرثومة في المعدة بعدد لا يُحصى من الطرق؛ مثل مصافحة شخص مصاب بالجرثومة. إن جانب التفكير السريع في عقولنا متصل بالقفز إلى الاستنتاجات السببية التي يمكنه التفكير فيها؛ مما يؤدي بنا عادة إلى اتخاذ قرارات خاطئة.
على العكس مما تقوله المعارف التقليدية، فإن ميل البشر نحو السببية لا يعمق من فهمنا للعالم من حولنا. في الكثير من الحالات، لا تتعدى كونها طريقًا إدراكيًّا مختصرًا يتسبب لنا بأفكار واهمة ولكنه يتركنا في الحقيقة غير مدركين لأي شيء مما يحدث في العالم من حولنا. مثلما كنا نستخدم العينات كطريق مختصرة لأننا لم نكن قادرين على معالجة جميع البيانات المتوفرة بين أيدينا، فإن إدراك السببية ما هو إلا طريق مختصرة تستخدمها عقولنا لتجنب بذل جهد كبير في التفكير يستغرق وقتًا.
في عالم البيانات المحدودة، كان إظهار مدى خطأ الإدراك السببي يستغرق وقتًا طويلاً. ولكن هذا في سبيله للتغير؛ ففي المستقبل، ستُستخدم العلاقات التبادلية في البيانات الضخمة من أجل إثبات أن المدركات السببية خاطئة، مظهرة وجود كم محدود، إن وجد، من العلاقات الإحصائية بين التأثير وسببه الافتراضي. إن أسلوب "التفكير السريع" موجود من أجل البحث عن الحقيقة بشدة وباستمرار.
ربما يدفعنا هذا الدرس للتفكير بعمق أكبر (وأبطأ) عندما نهدف لفهم العالم من حولنا. ولكن حتى أسلوب تفكيرنا البطيء ــ الطريقة الثانية التي نفحص بها السببية ــ سترى قواعدها تتغير بفضل العلاقات التبادلية للبيانات الضخمة.
في حياتنا اليومية، عادة ما نفكر بطرق سببية في أن ما نعتقد أنها أمور سببية من السهل توضيحها، ولكن الحقيقة ليست بهذه السهولة؛ فعلى النقيض من العلاقات التبادلية؛ حيث تكون المعادلات الرياضية مباشرة، لا توجد أية طريقة رياضية "لإثبات" السببية. لا يمكننا حتى أن نعبر عن العلاقات السببية بسهولة بمعادلات معيارية. إذن، حتى وإن فكرنا في الأمور ببطء وتروٍّ، فإن استنتاج وجود العلاقات السببية أمر في غاية الصعوبة. ولأن عقولنا قد اعتادت على عالم يفتقر للمعلومات، فإننا نُغرى بالتفكير دون وجود معلومات كافية، حتى وإن كانت هناك الكثير من العوامل جلية للعيان لتقلل من تأثير سبب بعينه.
خذ مثالاً للقاح ضد داء الكلب. في السادس من شهر يوليو عام 1885، حضر إلى الكيميائي الفرنسي "لويس باستير" "جوزيف مايستر" الفتى في التاسعة من عمره الذي عقره كلب مسعور. كان "باستير" قد ابتكر لقاحًا ونجح في تجاربه ضد داء الكلب. ألح والدا "مايستر" على "باستير" بأن يستخدم لقاحه لعلاج طفلهم. وفعل،

ونجا "جوزيف مايستر". وقالت الصحافة حينها إن "باستير" قد أنقذ الفتى من موت محقق ومؤلم.
ولكن هل فعل ذلك بالفعل؟ كما تبين فيما بعد أن واحدًا من بين كل سبعة أشخاص عقرتهم الكلاب المسعورة يقضي عليهم المرض. حتى مع الزعم بأن مصل "باستير" التجريبي كان فعالاً، هناك احتمال بنسبة 85% أن الصبي كان سينجو على أية حال.
في المثال السابق، تم الحكم بأن المصل قد عالج "جوزيف مايستر". ولكنَّ هناك صلتين سببيتين في هذه المسألة: الأولى بين المصل وفيروس داء الكلب، والثانية بين عضة الكلب المسعور والإصابة بالمرض. حتى وإن كان الأمر الأول صحيحًا فإن الثاني يكون صحيحًا في حالات معدودة.
لقد تغلب العلماء على تحدي تأكيد السببية من خلال التجارب؛ حيث إن السبب المفترض قد يتم تطبيقه أو إخفاؤه بعناية. إذا كانت التأثيرات تتوافق مع ما إذا كان السبب قد تم تطبيقه أم لا، فإنها تفترض وجود صلة سببية. مع زيادة التحكم بحرص في الظروف، زادت احتمالية صحة الصلة السببية التي حددتها.
إذن، مع التشابه الكبير بينها وبين العلاقات التبادلية، لا يمكن أن يتم إثبات السببية إلا نادرًا إن تمكن أحد من ذلك، بل تظهر مع وجود هامش كبير للاحتمالية. ولكن على العكس من العلاقات التبادلية، لا تكون التجارب التي تستنتج الصلات السببية عملية أو تطرح أسئلة أخلاقية صعبة. كيف يمكننا إجراء تجربة سببية من أجل تحديد السبب وراء نجاح بعض المصطلحات البحثية عبر الإنترنت في توقع انتشار فيروس الأنفلونزا؟ أما بالنسبة للمصل المضاد لداء الكلب، هل سنُخضع العشرات، أو ربما المئات من المرضى لهذا الموت المؤلم ـــ كجزء من "مجموعة التحكم" التي لم تأخذ المصل ـــ على الرغم من وجود المصل المضاد لمرضهم؟ حتى في المجالات التي تكون فيها التجارب فعالة، فإنها تظل مكلفة وتستغرق الكثير من الوقت.
مقارنة بهذا الأمر، التحليلات غير السببية، مثل العلاقات التبادلية، عادة ما تكون سريعة وغير مكلفة. على العكس من الصلات السببية، نمتلك المنهجيات الرياضية والإحصائية لتحليل العلاقات والأدوات الرقمية اللازمة لإثبات قوتها بثقة.
علاوة على ذلك، فإن العلاقات التبادلية ليست قيمة في كونها صحيحة فحسب، بل أيضًا تشير إلى الطريق نحو التحقيقات السببية, من خلال إخبارنا عن احتمالية وجود صلة بين أمرين، تمكننا من التحقيق في الأمر بصورة أعمق لمعرفة ما إذا كانت العلاقات السببية متواجدة أم لا، وإذا كانت متواجدة فما السبب وراء وجودها. إن هذه الآلية الرخيصة والسريعة للترشيح تقلل من تكلفة التحليل السببي من خلال تجارب خاصة متحكم بها. من خلال العلاقات المتبادلة يمكننا أن نحصل على نظرة خاطفة عن المتغيرات المهمة التي يمكننا استخدامها فيما بعد في التجارب لفحص السببية.
ولكن علينا أن نكون حذرين. إن العلاقات التبادلية تكون قوية، ليس لأنها تعرض إدراكًا أعمق فحسب؛ بل لأن الإدراك الذي تقدمه يكون واضحًا نسبيًّا. عادة ما يُشوش هذا الإدراك عندما تعود السببية للظهور. على سبيل المثال، شركة Kaggle،

الشركة التي تنظم مسابقة فحص البيانات بين الشركات المفتوحة أيًّا كانت لدخولها، نظمت مسابقة عام 2012 عن جودة السيارات المستعملة. قام أحد تجار السيارات المستعملة بتقديم بيانات عنها للإحصائيين المشاركين في المسابقة لإنشاء معادلة رياضية لتوقع أي من هذه السيارات ستكون جاهزة للبيع في مزاد علني في حين أن بها بعض المشكلات. أظهر تحليل العلاقات التبادلية أن السيارات المطلية باللون البرتقالي عادة ما تكون أقل عرضة لوجود عيوب بها ــ وذلك بنسبة تعادل حوالي نصف معدل متوسط السيارات الأخرى.

حتى وإن قرأنا هذا الأمر، فقد فكرنا بالفعل في سبب حدوثه: هل يحتمل أن يكون الأشخاص الذين اشتروا السيارات البرتقالية لا يحبون السيارات وأولوا كامل اهتمامهم للدراجات الهوائية؟ هل يعني اللون ذاته أن هذه السيارات قد تم صنعها بطريقة أكثر حرصًا وتخصيصًا؟ أو ربما لأن السيارات البرتقالية تكون ملحوظة بدرجة كبيرة على الطرق ولذلك لا تكون طرفًا في الكثير من الحوادث؛ لذا تكون في حالة أفضل من غيرها عندما تباع مستعملة؟

سرعان ما وقعنا في شرك شبكة من الفرضيات السببية المتنافسة فيما بينها. ولكن محاولتنا لكشف الأمور بهذه الطريقة تجعل منها أكثر غموضًا. إن العلاقات التبادلية موجودة، ويمكننا إثبات وجودها رياضيًّا. ولكننا لا يمكننا أن نفعل المثل مع العلاقات السببية بسهولة؛ لذا سيكون من الأفضل أن نتوقف عن محاولة شرح السبب وراء العلاقات التبادلية: المشكلة بدلاً من سببها. وإلا يجب علينا أن ننصح أصحاب السيارات بطلاء سياراتهم باللون البرتقالي حتى تقل عيوب محركاتها ــ فكرة سخيفة.

بوضع هذه الحقائق في الاعتبار، من المفهوم أن تحليل العلاقات التبادلية وما يماثلها من المنهجيات التي تعتمد على البيانات الصعبة تكون في مقدمة المنهجيات التي تحتوي على الصلات السببية، نتيجة "للتفكير السريع". ولكن في عدد متزايد من السياقات، تفيد مثل هذه التحليلات وتكون فعالة أكثر من التفكير البطيء السببي الذي تم تلخيصه من خلال التجارب المتحكم بها بعناية (عالية التكلفة والتي تستغرق الكثير من الوقت).

في الأعوام القليلة السابقة، حاول العملاء التقليل من تكلفة المناهج التجريبية من أجل فحص المسببات، على سبيل المثال، من خلال المزج بمهارة بين استقصاءات الرأي لإنشاء "تجارب تقريبية". قد يعمل هذا الأمر على جعل التحقيقات السببية أسهل، ولكن ميزة كفاءة المنهجيات السببية من الصعب التغلب عليها. علاوة على ذلك، فإن البيانات الضخمة نفسها تساعد المسائل السببية؛ حيث إنها تقود الخبراء نحو المسببات المرجحة ليقوموا بالتحقيق فيها. في العديد من الحالات، يتم بحث أكثر عمقًا عن السببية بعد أن تقوم البيانات الضخمة بعملها، عندما نرغب في البحث عن السبب، دون أن نقدر المشكلة وحدها.

لا يمكن تجاهل السببية، ولكن تم التغلب على أسسها؛ حيث إنها المكون الرئيسي للمعنى. إن البيانات الضخمة تساعد التحليلات غير السببية بشكل كبير؛ حيث إنها

عادة ما تستبدل الأبحاث السببية بها. إن لغز انفجار فتحات الصرف في مانهاتن خير مثال على هذا الأمر.

الإنسان مقابل فتحات الصرف

كل عام، تُغطى عدة مئات من فتحات الصرف في مدينة نيويورك بالدخان بسبب اشتعال النيران داخلها. تنفجر أحيانًا أغطية فتحات الصرف الحديدية التي تزن 300 رطل في الهواء بارتفاع عدة طوابق قبل أن تصطدم بالأرض، ويا له من أمر سيئ.

تقوم شركة "كون إديسون"، شركة الخدمات العامة المختصة بصيانة الكهرباء بالمدينة، بإجراء فحوصات وصيانة دورية لفتحات الصرف بالمدينة كل عام. في الماضي، كانت الشركة تعتمد بشكل رئيسي على الصدفة، على أمل أن تكون فتحات الصرف التي وضعتها قيد الصيانة من بين فتحات الصرف المعرضة للانفجار. وكان الأمر أفضل قليلاً من السير العشوائي لفحص فتحات الصرف في شارع وول ستريت. وفي عام 2007، تحولت الشركة لطلب المساعدة من الإحصائيين بجامعة كولومبيا على أمل استخدام البيانات التاريخية لشبكة الصرف، مثل المشكلات السابقة وما الذي تتصل به البنى التحتية، لتوقع أي فتحات الصرف من المحتمل أن تقع بها المشكلات، ومن ثم تعلم الشركة أين تجند جميع مواردها.

كان الأمر عبارة عن معضلة بيانات ضخمة معقدة؛ فقد كان هناك حوالي 94 ألف ميل من الكابلات تحت الأرض في مدينة نيويورك، والتي تكفي لتلتف حول الكرة الأرضية ثلاث مرات ونصف. يوجد في منهاتن وحدها حوالي 51 ألف فتحة صرف وصندوق خدمة. بعض هذه البنى التحتية يعود تاريخ إنشائها إلى عصر "توماس إديسون"، الذي سُميت الشركة تيمنًا به. كان واحدًا من بين كل 20 كابلًا قد تم تركيبه قبل عام 1930. على الرغم من أن سجلات البنى التحتية قد بدأت في ثمانينات القرن التاسع عشر، فإنها كانت موزعة على عدد كبير من الوثائق ـــ ولم تكن معدة لتحليل البيانات، فقد أعدتها أقسام المحاسبة أو عمال الطوارئ الذين كانوا يكتبونها بخط اليد في صورة "بطاقات المشكلات". إن القول بأن البيانات كانت في حالة شديدة من الفوضى كان تقليلاً لما عليه الأمر بالفعل. هناك أحد الأمثلة التي أبلغ بها الإحصائيون وهو مصطلح "صندوق الخدمة"، وهو جزء شائع من البنى التحتية، يشتمل على حوالي 38 متغيرًا من بينها SB، S، S/B، S.B، S?B، S.B.، SBX، S/BX، SB/X، S/XB، S.BX، S&BX، S?BX، S BX، S/B/X، S BOX، SVBX، SERV BX، SERV-BOX، SERV/BOX , SERVICE BOX.

وكان على المعادلة الحاسوبية أن تستوعب جميع هذه المتغيرات.

تتذكر "كينثيا رودين" الإحصائية والمدققة البيانية التي تعمل حاليًّا في معهد ماساتشوسيتس للتكنولوجيا والتي كانت تقود الفريق، قائلة: "لقد كانت البيانات بدائية بشكل مذهل. لقد حصلت على نسخة مطبوعة من جميع جداول الكابلات المختلفة. إذا قمت بطباعة هذه الجداول، فلن يمكنك أن تحملها بين يديك دون أن تجرها وراءك على الأرض. وعليك أن تخرج بأمر مفهوم من كل هذه البيانات ـــ أن

تبحث داخلها عن المعلومات المفيدة، ومهما تكلف الأمر من أجل أن تحصل على نموذج توقع جيد".

لكي يبدءوا العمل، كان على "رودين" وفريقها أن يستخدموا جميع البيانات المتوفرة بين أيديهم ليس مجرد عينة منها؛ حيث إن أي واحدة من عشرات الآلاف من فتحات الصرف قد تكون قنبلة موقوتة؛ لذا كان عليهم استخدام معادلة ن=الجميع. وعلى الرغم من أن الخروج بأسباب حدوث هذا الأمر كان يكفي، فإنه كان سيمر قرن من الزمان ولا تزال النتائج خاطئة أو غير كاملة. إن الأمر الأفضل لإتمام هذه المهمة هو التوصل للعلاقات التبادلية داخل تلك البيانات. كانت "رودين" لا تُعنى كثيرًا بالسبب بل كانت تُعنى بالمشكلة ذاتها ــ على الرغم من أنها كانت تعلم عندما جلس الفريق مع المديرين التنفيذيين لشركة "كون إديسون" بأن المهووسين بالإحصاء سيعدلون من الوضع من أجل تقييماتهم. قد تصدر التوقعات من خلال آلة، ولكن العملاء كانوا من البشر، ويميل البشر لمعرفة الأسباب ليفهموا الأمر.

كشف الفحص المتعمق للبيانات عن المعلومات القيمة التي كانت تأمل "رودين" في العثور عليها. بعد تنظيم البيانات الفوضوية بحيث تتمكن الآلات من معالجتها، بدأ الفريق بالتنبؤ بحدوث 106 من كوارث فتحات الصرف الكبرى، ثم قلصوا القائمة لتشمل بضعة مؤشرات من أقوى المؤشرات على حدوث المشكلة. خلال اختبار للبنية التحتية لمصادر الطاقة في حي برونكس، قاموا بتحليل جميع البيانات التي توافرت لهم حتى منتصف عام 2008، ثم استخدموا هذه البيانات لتوقع مواقع المشكلات لعام 2009. وقد نجح الأمر نجاحًا باهرًا. كانت أول 10% من فتحات الصرف المدرجة في قائمتهم تحتوي على 44% من فتحات الصرف التي انتهى بها الأمر وحدثت بها حوادث جسيمة.

في نهاية الأمر، كانت أهم العوامل هي قدم الكابلات وما إذا كانت فتحات الصرف قد مرت بمشكلات سابقة أم لا. كان هذا الأمر مفيدًا للغاية، فكما تبين، تمكنت مقاربة شركة "كون إديسون" من تحديد قاعدة لتقييم الأمر بسهولة. القدم والمشكلات السابقة؟ ألا يبدو هذا الأمر واضحًا؟ حسنًا، نعم ولا. فمن جانب، كما يحب واضع نظريات الشبكات "دانكن واتس" أن يقول: "يتضح كل شيء بمجرد أن تعلم إجابته". (عنوان أحد كتبه). ومن جانب آخر، من المهم أن نتذكر أنه كان يوجد 1 من التوقعات في بداية النموذج. لم يكن واضحًا كيفية تقييمها ومن ثم تحديد أولوية عشرات الآلاف من فتحات الصرف، كل واحدة من عشرات الآلاف من المتغيرات التي أضيفت إلى ملايين النقاط البيانية الموجودة بالفعل ــ ولم تكن البيانات نفسها في حالة تسمح بمعالجتها.

كانت حالة فتحات الصرف المنفجرة توضح فكرة أن البيانات قد تم استخدامها بشكل جديد من أجل علاج المشكلات الواقعية. ومن أجل تحقيق هذا الأمر، كنا بحاجة إلى تغيير الطريقة التي نعمل بها. كان علينا استخدام جميع البيانات المتوفرة بين أيدينا، وكذلك جمع كل ما أمكننا جمعه منها، وليس فقط جزءًا صغيرًا منها. كنا بحاجة إلى تقبل فوضى البيانات بدلاً من معاملة الدقة على أنها الأولوية القصوى

من أجل معالجتها. وكان علينا أن نثق بشدة في العلاقات التبادلية دون أن ندرك إدراكًا كاملاً القاعدة السببية للتوقعات.

هل هذه نهاية المنهج النظري؟

إن البيانات الضخمة تغير من طريقة استيعابنا واستكشافنا للعالم من حولنا. في عالم البيانات المحدودة، كانت المناهج التجريبية هي التي تقودنا لإدراك الطريقة التي يدور بها العالم من حولنا، والتي حاولنا فيما بعد أن نتيقن من صحتها من خلال جمع وتحليل البيانات. في المستقبل، سوف يعتمد استيعابنا للعالم من حولنا أكثر على البيانات الوفيرة بدلاً من المناهج التجريبية.

كانت هذه المناهج التجريبية مشتقة من نظريات علم الاجتماع، والذي ساعد بدوره على تفسير و/أو توقع العالم من حولنا. في أثناء تحولنا من العالم الذي يعتمد على المناهج التجريبية إلى العالم الذي يعتمد على البيانات، قد نرغب في التفكير في أننا لم نعد بحاجة إلى النظريات.

في عام 2008، أعلن رئيس تحرير مجلة *Wired* "كريس أندرسون": "إن وفرة البيانات جعلت من المنهج العلمي أمرًا عتيقًا". في موضوعه الذي أشير إليه على غلاف أحد أعداد المجلة تحت عنوان "عصر البيتابايت"، ادعى "أندرسون" أن الأمر قد وصل إلى مرحلة "نهاية المنهج النظري". إن الآلية التقليدية للاكتشافات العلمية ـــ التي تعتمد على المنهج التجريبي الذي يتم اختباره مقابل الواقع باستخدام نموذج للمسببات الخفية ـــ في طريقها للاندثار، كما يقول "أندرسون"، وسيتم استبدالها بالتحليل الإحصائي للعلاقات التبادلية الخالصة المجردة من النظريات.

من أجل دعم ما قال، شرح "أندرسون" كيف أصبحت الفيزياء الكمية مجالاً نظريًّا بحتًا، بسبب أن التجارب كانت مكلفة للغاية ومعقدة للغاية وأكبر من إمكانية القيام بها. يقول أيضًا إن هناك منهجًا نظريًّا لم يعد يصلح لتفسير الواقع، وضرب أمثلة على الأسلوب الجديد في التحليل وهي محرك البحث من شركة جوجل وترتيب الحمض النووي البشري؛ حيث كتب قائلاً: "إننا نحيا في عالم تحل فيه الكميات الضخمة من البيانات والرياضة التطبيقية محل جميع الأدوات الأخرى التي قد نهدف إلى تحقيقها. مع وجود كَمٍّ كافٍ من البيانات، تتحدث الأرقام عن نفسها؛ حيث تسمح لنا المساحات التخزينية الكبيرة بأن نقول: إن العلاقات التبادلية كافية".

لقد أطلق المقال العنان لجدل محتدم ومهم، على الرغم من أن "أندرسون" قد تراجع عن هذه الادعاءات الجريئة. ولكن يستحق ما قاله الفحص والتدقيق. في الحقيقة، ظل "أندرسون" يؤكد على ما كتبه حتى وقت قريب؛ حيث إننا عندما نهدف إلى تحليل وفهم العالم من حولنا، فإننا نحتاج إلى نظريات لاختبارها. ولكن على النقيض، في عصر البيانات الضخمة، يستمر الجدل بأننا لا نحتاج إلى النظريات: يمكننا فقط أن نفحص البيانات. وفي المستقبل، قد يفترض هذا الأمر أن جميع القواعد الشاملة عن كيفية دوران العالم من حولنا، كيف يتصرف البشر، وما الذي يشتريه المستهلكون، ومتى تتعطل قطع الغيار، وغيرها قد تكون غير ذات صلة بالنظريات حيث ستهيمن البيانات الضخمة على مجريات الأمور.

قد تتضمن مقولة "نهاية المنهج النظري" أنه على الرغم من وجود النظريات في مجالات واقعية مثل الفيزياء أو الكيمياء، فإن تحليل البيانات الضخمة لا يحتاج إلى أية نماذج مفاهيمية، وهو أمر منافٍ للعقل.

إن البيانات الضخمة ذاتها يمكن إيجادها داخل النظريات. على سبيل المثال، تقوم بتوظيف النظريات الإحصائية والرياضية، وفي أحيان أخرى تستخدم نظريات علوم الحاسب أيضًا. نعم إنها ليست نظريات عن سببية عمل ظاهرة معينة مثل الجاذبية، إلا أنها تُعتبر نظريات رغمًا عن ذلك. وكما وضحنا من قبل، فإن النماذج المعتمدة على هذه النظريات تمتلك قدرة توقع كبيرة. في حقيقة الأمر، قد تقدم البيانات الضخمة نظرة جديدة وأفكارًا جديدة لأنها تكون غير مثقلة بأساليب التفكير التقليدية والتوجهات الموروثة التي تظهر في النظريات المختصة بمجال محدد.

علاوة على ذلك، ولأن تحليل البيانات الضخمة يعتمد على النظريات، لا يمكننا أن نتجنبها؛ حيث إنها ما يشكل كلًّا من أساليبنا ونتائجنا؛ حيث يبدأ الأمر بكيفية اختيارنا للبيانات. قد تعتمد قراراتنا على الظروف المناسبة: هل تتوفر البيانات بشكل جيد؟ أو بالاقتصاديات: هل يمكن الحصول على البيانات بتكلفة أقل؟ إن خياراتنا تتأثر بالنظريات. إن ما نختاره يؤثر على ما نجده كما تقول الباحثتان في مجال التكنولوجيا الرقمية "دانا بويد" و"كايت كراوفورد". في النهاية نقول إن شركة جوجل قد استخدمت المصطلحات البحثية كدليل للتوصل إلى انتشار فيروس الأنفلونزا، وليس طول شعر الناس. وبالمثل، عندما نقوم بتحليل البيانات، فإننا نختار الأدوات التي تعتمد على النظريات. وعندما نتمكن من التوصل إلى نتائج نقوم مرة أخرى بتطبيق النظريات. من الواضح أن عصر البيانات الضخمة لا يخلو من النظريات؛ حيث إنها تتواجد في جميع ما يتعلق بالبيانات الضخمة.

يستحق "أندرسون" التقدير لأنه طرح السؤال المناسب, ولأنه قام بهذا قبل أن يقوم به أي شخص آخر. قد لا تتسبب البيانات الضخمة في الوصول إلى مرحلة "نهاية المنهج النظري"، ولكنها قامت بلا شك بتغيير الطريقة التي ننظر بها إلى العالم من حولنا. سيحتاج هذا التغيير إلى كثير من الوقت حتى نعتاد عليه؛ حيث إنه يتحدى الكثير من القواعد الثابتة. ولكن القيمة التي أطلقت لها العنان لن تعتبر مجرد تحول كبير فحسب، بل تحول حتميّ أيضًا.

قبل أن نصل إلى تلك المرحلة، لا يهم كيف وصلنا إليها. يحب الكثير من العاملين في صناعة التكنولوجيا أن يثنوا على التحول إلى استخدام الأدوات الرقمية الحديثة، من الرقاقات السريعة إلى البرمجيات الفعالة؛ لأنهم هم صناع هذه الأدوات. إن الابتكار التقني أمر مهم، ولكنه ليس على قدر الأهمية الذي قد يعتقده المرء. إن السبب الأعمق لجميع هذه التوجهات هو أننا نملك الكثير من البيانات، والسبب في امتلاكنا الكثير من البيانات هو أننا نستخلص ونؤول المزيد من الأوجه المختلفة للواقع بشكل يتعلق بالبيانات، وهذا هو موضوع الفصل التالي.

* متوافر لدى مكتبة جرير.

5
التحول إلى البيانات

كان "ماثيو فونتاين موراي " الضابط الواعد بالبحرية الأمريكية متجهًا لتولي مهمة جديدة على البارجة "**كونسورت**" عام 1839، عندما انزلقت مركبته عن الطريق وانقلبت رأسًا على عقب، وطار منها في الهواء، وكانت صدمته بالأرض قوية لدرجة كسر عظمة فخذه وخلع ركبته. قام أحد الأطباء المحليين بإعادة المفصل إلى مكانه ولكن كانت عظام الفخذ قد تم تجبيرها بطريقة خاطئة وكانت بحاجة إلى كسرها مرة أخرى بعد بضعة أيام. تركت تلك الإصابة "موراي " وهو في الثالثة والثلاثين من العمر معاقًا جزئيًّا وغير صـالح للعمل في البحر. بعد مرور ثلاث سنوات من النقاهة، وضعته قوات البحرية الأمريكية في عمل مكتبي، في وظيفة مملة كرئيس لمخزن الخرائط والوثائق.

تبين فيما بعد أن هذا المكان هو المكان المثالي بالنسبة له للعمل به؛ حيث إنه كملاح شاب، كان متحيرًا من الطريقة التي كانت تسير بها السفن في البحر بطريقة متعرجة بدلاً من اتخاذ طرق مباشرة. عندما استطلع آراء قباطنة السفن عن هذا الأمر، ردوا عليه بأنه من الأفضل الإبحار في مسار معروف بدلاً من المخاطرة بالإبحار في مسارات غير معروفة والتي قد تخبئ لهم مخاطر لا يعلمها إلا الله وحده. كانوا ينظرون للمحيط على أنه مملكة لا يمكن توقع أفعالها؛ حيث يواجه البحارة الأمور غير المتوقعة مع كل ريح وموج يضرب جوانب سفنهم.

ولكن من الرحلات التي شارك بها، كان "موراي " يعلم أن هذا الأمر ليس صحيحًا؛ فقد كان يرى أنماطًا في كل مكان من المحيط. خلال فترة توقف طويلة في فالبارايسو بدولة تشيلي، لاحظ أن الرياح منتظمة كالساعة؛ حيث تتوقف العواصف في فترة ما بعد الظهيرة لتتحول إلى نسيم عليل، كما لو أن شخصًا ما أغلق الصنبور. في رحلة أخرى، عبر المياه الزرقاء الدافئة لتيار الخليج في أثناء تدفقها بين الجدران الداكنة لمياه المحيط الأطلنطي المالحة، والتي كانت تتميز عن المياه من حولها وذات مسار ثابت كما لو كانت نهر الميسيسيبي. لا شك في أن البرتغاليين كانوا يبحرون في المحيط الأطلنطي طوال قرون معتمدين على رياح شرقية وغربية ثابتة يُطلق عليها "الرياح التجارية" (والتي كان يُطلق عليها الإنجليز القدامى "الطريق" أو "المسار" والتي أصبحت فيما بعد مرتبطة بالتجارة).

عندما كان يصل الضابط البحري "موراي " إلى ميناء جديد، كان يبحث عن قباطنة البحار القدامى من أجل الحصول منهم على المعرفة بناءً على الخبرات التي تناقلتها الأجيال. عرف الكثير عن المد والجزر والرياح والتيارات البحرية التي كانت لا تتغير ــ والتي لا مكان لها في الكتب التي تصدرها البحرية الأمريكية لتوزعها على بحارتها. كانت البحرية تعتمد بدلاً من ذلك على الخرائط التي قد يعود تاريخ إعدادها إلى مائة عام مضت في بعض الأحيان مع وجود حذف أو أخطاء كثيرة بها. ومن

موقعه الجديد كمشرف على مخزن الخرائط والوثائق، كان هدفه أن يصلح هذا الأمر.

بعدما تسلم وظيفته، كان يحتفظ بأجهزة الباروميتر والبوصلات وآلات السدس والكرونوميتر ضمن الأشياء التي كانت محفوظة بالمخزن. ولاحظ أيضًا وجود عشرات الآلاف من الكتب والخرائط والرسوم البيانية البحرية. وجد أيضًا صناديق عتيقة مليئة بسجلات سفن قديمة لقباطنة قدامى من سلاح البحرية الأمريكية. كان من سبقوه في العمل بسلاح البحرية الأمريكية يعتبرون الكتب التي تمدهم بها البحرية مجرد هراء؛ حيث وضعوا على هوامش السجلات جملًا أو رسومات ساخرة، وهو ما جعلها تبدو أحيانًا أكثر من مجرد سجلات لأحوال السفينة، بل طريقة للهروب من الملل الذي يشعرون به.

وبينما كان "موراي " يقوم بإزالة الأتربة عن السجلات التي تحمل بقع الماء المالح ويبحث في داخلها، أصبح متشوقًا للغاية لمعرفة كل ما تحتويه. لقد وجد المعلومات التي كان بحاجة إليها: سجلات عن الرياح والمياه والمناخ في أماكن بعينها وفي تواريخ محددة. وعلى الرغم من أن بعض هذه السجلات لم تكن تحوي أمورًا ذات قيمة، فإن الكثير منها كان مكدسًا بالمعلومات المفيدة. قام "موراي " بجمع كل هذه المعلومات معًا وأصبح من الممكن عمل نوع جديد من الخرائط الملاحية. بدأ "موراي " ودستة "حاسباته" ــ المسمى الوظيفي الذي كان يطلق حينها على من يقوموا بحساب البيانات ــ العملية الشاقة لاستخراج وجدولة المعلومات الموجودة داخل السجلات التالفة.

جمع "موراي " جميع البيانات وقسم المحيط الأطلنطي بأكمله إلى أقسام ذات خمس درجات للطول والعرض لكل منها. وراقب درجات الحرارة وسرعة واتجاه الرياح والأمواج لكل قسم على مدار الشهر؛ حيث إن هذه الظروف تتغير بتغير أوقات العام. عندما تم مزج البيانات مع بعضها البعض، كشفت عن أنماط وأشارت إلى وجود طرق إبحار أكثر فاعلية.

تسببت معارف أجيال من البحارة في توجيه السفن إما إلى المياه الهادئة أو إدخالهم في عمق الرياح والتيارات البحرية العاتية. في أحد الطرق البحرية الشائعة من نيويورك إلى ريو دي جانيرو، كان البحارة يميلون أكثر إلى مقاومة الطبيعة بدلاً من الاعتماد عليها؛ حيث تدرب قباطنة السفن الأمريكان على تجنب مخاطر الإبحار المباشر نحو ريو دي جانيرو؛ لذا كانت سفنهم تبحر في اتجاه الجنوب الشرقي قبل أن يتحول اتجاهها نحو الجنوب الغربي بعد أن تعبر خط الاستواء. كانت المسافة التي يبحرونها تساوي عبور المحيط الأطلنطي ثلاث مرات كاملة. تبين أن هذا الطريق الملتف حول نفسه عبارة عن هراء؛ حيث إن الإبحار نحو الجنوب بشكل مباشر كان أمرًا جيدًا.

لكي يحسن من دقة نتائجه، كان "موراي " يحتاج إلى المزيد من المعلومات. لذا قام بتفعيل نظام معياريّ لتسجيل بيانات السفن وتعميمه على جميع سفن البحرية الأمريكية وتسليمه عند الوصول إلى المواني. كانت السفن التجارية في أمس الحاجة لخرائط "موراي "؛ لذا أصر "موراي " على أن تقوم هذه السفن هي الأخرى

بتسليم سجلات رحلاتها إليه (نسخة قديمة من شبكات التواصل الاجتماعي الفيروسية). أعلن "موراي " قائلاً: "إن كل سفينة تبحر في عرض البحر ـــ يمكن اعتبارها مرصدًا عائمًا ـــ محراب للعلم". من أجل زيادة دقة الخرائط، سعى "موراي " للحصول على نقاط بيانية أخرى (كما فعلت شركة جوجل عند اعتمادها على معادلة PageRank لتشتمل على المزيد من المؤشرات). طلب من القباطنة أن يلقوا بزجاجات تحتوي على ملاحظات مكتوب بها تاريخ اليوم والموقع والرياح والتيار البحري السائد في البحر على فترات منتظمة، وأن يحصلوا على أية زجاجة من هذه الزجاجات قد يرونها. رفعت الكثير من السفن علمًا يدل على أنها مشاركة في مشروع تبادل المعلومات (وكان يحمل شعارات مشاركة الروابط الذي يظهر على بعض الصفحات الإلكترونية في الوقت الحالي).

من خلال البيانات، أعلنت الطرق الملاحية عن نفسها؛ حيث كانت الرياح والتيارات البحرية مواتية. قصرت خرائط "موراي " من الرحلات الطويلة، تقريبًا بمقدار ثلث الوقت؛ مما وفر على السفن التجارية الكثير من المال. كتب أحد القباطنة الممتنين من عمل "موراي " قائلاً: "قبل أن أستخدم عملك كنت أجول في البحر جيئة وذهابًا معصوب العينين". حتى البحارة القدامى الذين رفضوا العمل بالنظام الجديد واعتمدوا على الطرق التقليدية أو على حدسهم كانوا يخدمون المسعى: عندما كانت رحلاتهم تستغرق وقتًا أكثر من اللازم أو يواجهون كارثة، كانوا يثبتون فائدة نظام "موراي ". بحلول عام 1855، عندما نشر عمله الرائع *The Physical Geography of the Sea*، كان "موراي " قد فحص 1.2 مليون نقطة بيانية. كتب "موراي ": "إن البحار الجديد بدلاً من تلمس طريقه في الظلام حتى تصل إليه أضواء الخبرة... سيجد في هذا النظام، وعلى الفور، أنه يحتوي على خبرات آلاف الملاحين لترشده على الطريق وسط المحيط".

كان عمله أساسيًّا عند وضع أول كابل تلغراف بحري يعبر المحيط الأطلنطي. وبعد حدوث اصطدام مأساوي في عرض البحر، قام "موراي " بسرعة شديدة بابتكار نظام الخطوط البحرية المتعارف عليه حتى وقتنا الحالي. لقد قام حتى بتطبيق أسلوبه على علم الفلك: عندما تم اكتشاف كوكب نبتون عام 1846، طرأت على "موراي " فكرة عبقرية بأن يتم فحص السجلات من أجل البحث عن الإشارات الخاطئة لهذا الكوكب على أنه نجم، الأمر الذي جعل العلماء يفحصون مداره جيدًا.

تم تجاهل "موراي " بشكل كبير في كتب التاريخ الأمريكي، ربما لأن هذا الشخص المنحدر من فيرجينيا قد استقال من بحرية (الاتحاد) في أثناء الحرب الأهلية وخدم كجاسوس في إنجلترا لصالح الكونفدرالية. ولكن قبل هذا بعدة سنوات عندما وصل إلى أوروبا ليحصل على الدعم من أجل خرائطه، نصبته أربع دول في منصب فارس وحصل على الوسام الفضي من ثماني دول أخرى من بينها الفاتيكان. وحتى مطلع القرن الحادي والعشرين، كانت خرائط الطيارين التي تصدرها البحرية الأمريكية لا تزال تحمل اسمه.

كان ضابط البحرية "موراي ، مستكشف البحار"، من بين أوائل من أدركوا وجود

قيمة خاصة في الكيان الضخم للبيانات والتي لا توجد في الكميات الأصغر حجمًا ـــ جوهر عقيدة البيانات الضخمة. الأمر الأهم، أنه أدرك أن سجلات السفن المتربة تلك تحتوي على "بيانات" يمكن استخراجها وجدولتها. وبمجرد أن فعل هذا الأمر، أصبح من رواد التحول إلى البيانات في العالم، ومن رواد الكشف عن البيانات من مواد لم يعتقد أي شخص أنها تحمل أية قيمة. مثله مثل "أورين إتزيوني" في شركة ، الذي استخدم أسعار تذاكر الطيران القديمة لينشئ تجارة رابحة، أو مثل مهندسي شركة جوجل الذين لو طبقوا طلبات البحث القديمة لفهموا انتشار فيروس الأنفلونزا، أخذ "موراي " المعلومات التي تم إنتاجها من أجل أمر ما وحولها إلى أمر مختلف تمامًا.

إن أسلوبه، الذي يشبه إلى حدٍّ كبير تقنيات البيانات الضخمة الحالية، كان مذهلاً مع الوضع في الاعتبار أنه تم بورقة وقلم رصاص. إن قصته توضح الدرجة التي وصل إليها استخدام البيانات في الفترات السابقة للرقمنة. اليوم، نميل إلى الدمج بين العصرين، ولكن من المهم جدًّا الحفاظ على استقلالية كل منهما. للحصول على فكرة أكثر شمولاً عن كيفية استخراج البيانات من الأماكن غير المتوقعة. فكر في أمثلة أكثر حداثة.

إن تقدير الأجزاء الخلفية من البشر هي فن وعلم يُسمى شيجيومي كوشيميزو، وهو على اسم أستاذ في المعهد الياباني المتطور لتكنولوجيا الصناعة بطوكيو. قلة فقط من يقولون إن الطريقة التي يجلس بها الأشخاص تحتوي على الكثير من المعلومات عنهم، ولكنها تحتوي على الكثير من المعلومات بالفعل. عندما يجلس شخص ما، فإن محيط الجسم الخارجي ووضعية الجلوس وتوزيع الوزن معلومات يمكن جمعها وجدولتها. قام "كوشيميزو" وفريقه بتحويل الأجزاء الخلفية من جسم الإنسان إلى بيانات من خلال قياس وضعية الجلوس من جميع الزوايا باستخدام حساسات مثبتة في مقاعد السيارات وتصنف كل نقطة بيانية على مقياس من الصفر وحتى 256. كانت النتيجة وجود شفرة رقمية فريدة لكل شخص. خلال التجربة، كان النظام قادرًا على التمييز بين مجموعة من الأشخاص بدقة تبلغ 98%.

لم يكن هذا البحث يُجرى هباءً، بل كان يُجرى من أجل تطوير أنظمة الحماية من سرقة السيارات. ستتمكن السيارة التي تم تركيب هذا النظام بها من إدراك ما إذا كان من يجلس خلف عجلة القيادة سائق آخر غير صاحب السيارة وستطلب السيارة رقمًا سريًّا من أجل استكمال قيادتها أو ربما توقف عمل المحرك. إن تحويل أوضاع الجلوس إلى بيانات يخلق خدمة قابلة للتطبيق وربما تجارة رابحة، وقد تتعدى استخداماتها مجرد حماية السيارات من السرقة. على سبيل المثال، يمكن للبيانات المجمعة أن تكشف عن أدلة تربط بين وضعية جلوس السائق والأمان على الطريق، مثل مراقبة تغير أوضاع جلوس السائق قبل الحوادث. يمكن للنظام أيضًا أن يشعر بنعاس السائق بسبب الإجهاد ويرسل تنبيهًا للسائق أو يُفَعِّل مكابح السيارة. لن يعمل هذا النظام على حماية السيارات من السرقة فحسب، بل سيتعرف أيضًا على وضعية جلوس اللص أيضًا (كما يقال).

أخذ البروفيسور "كوشيميزو" أمرًا لم تتم معاملته أبدًا على أنه بيانات ـــ أو لم

يتخيل أحد أنه يحتوي على أية قيمة معلوماتية ـــ وتحوله إلى الصورة الرقمية المحددة، مثلما فعل الضابط البحري "موراي " عندما أخذ المواد التي بدا أنها غير ذات فائدة واستخرج منها المعلومات وحولها إلى بيانات مفيدة. وبفعل هذا الأمر سمح باستخدام البيانات بطريقة مستحدثة وبأن تخلق قيمة فريدة.

إن كلمة "بيانات" تعني باللاتينية كلمة "مُعطَى"، أي تحمل إيحاء بمفهوم "الحقيقة". أصبحت البيانات عنوانًا لعمل كلاسيكي من تأليف "إقليدس"، والذي يشرح فيه الهندسة من وجهة النظر المعروفة أو التي يمكن إظهارها لتكون معروفة. في الوقت الحالي، تشير البيانات إلى وصف أمر ما يسمح لها بأن تُخزن وتُعالج وتُنظم. لا يوجد مصطلح جيد بعد لأنواع التحويل التي أنتجها الضابط البحري "موراي " والبروفيسور "كوشيميزو"؛ لذا دعنا نُطلق عليها "التحول إلى البيانات". من أجل تحليل بيانات أية ظاهرة هي أن نضعها في صورة محددة بحيث يمكن جدولتها وتحليلها.

مرة أخرى، أقول إن هذا الأمر بعيد كل البعد عن الرقمنة، وهي العملية التي يتم فيها تحويل المعلومات التناظرية إلى أصفار وآحاد في شفرة ثنائية بحيث يمكن للحواسب فهمها. إن الرقمنة ليست أول الأمور التي فعلناها مع الحواسب؛ فقد كانت الحقبة الأولى من ثورة الحواسب حسابية، كما يدل عليها لفظ الكلمة ذاته. كنا نستخدم الآلات لتقوم بالعمليات الحسابية التي كانت تستغرق وقتًا طويلاً بالطرق السابقة: جداول مسارات الصواريخ، إحصاءات عدد السكان، وحالة الطقس. بعد ذلك جاء العصر الرقمي وتم تحويل كل المحتوى التناظري إلى رقمي؛ لذا عندما نشر "نيكولاس نيجروبونتي" من معمل الإعلام بمعهد ماساتشوسيتس للتكنولوجيا، كتابه الرائع عام 1995 تحت عنوان *Being Digital*، كان من بين أفكاره الرئيسية تحويل الذرات إلى بايتات. قمنا برقمنة المحتوى النصي على نطاق واسع في تسعينيات القرن العشرين. ومؤخرًا، مع زيادة القدرة التخزينية للمعلومات، وقوة المعالجة، وسرعة الإنترنت، قمنا بفعل المثل مع أنواع أخرى من المحتويات أيضًا، مثل الصور ومقاطع الفيديو والموسيقى.

في الوقت الحالي، هناك اعتقاد ضمني بين جموع التقنيين بأن البيانات الضخمة تعود أصولها إلى ثورة السليكون، ولكن هذا الأمر ليس صحيحًا. لا شك في أن أنظمة تكنولوجيا المعلومات الحديثة جعلت من البيانات الضخمة في متناولنا، ولكن فيما يتعلق بأصولها فإنها تعود إلى سعي الجنس البشري المستمر لقياس وتسجيل وتحليل العالم. إن ثورة تكنولوجيا المعلومات تظهر في كل ما حولنا، ولكن كان كل التركيز فيما مضى على التكنولوجيا فقط، وحان الوقت الآن لنصب تركيزنا على المعلومات.

من أجل الحصول على معلومات قابلة للقياس، أو الحصول على البيانات، علينا أن نعرف أولاً كيفية قياسها وكيفية تسجيل ما نقيسه. يحتاج هذا الأمر إلى الأدوات المناسبة، وكذلك ضرورة وجود الرغبة في الحساب والتسجيل. هذان الأمران كلاهما من ضروريات التحول إلى البيانات، ولقد وضعنا أحجار الأساس اللازمة للتحول إلى البيانات قبل قرون من بداية العصر الرقمي.

قياس العالم

إن القدرة على تسجيل المعلومات تعتبر من الخطوط الفاصلة بين المجتمعات البدائية والمتقدمة. إن القياسات الأساسية للطول والوزن كانت من بين أقدم الأدوات المفاهيمية في الحضارات القديمة. بحلول الألفية الثالثة قبل الميلاد، تطورت فكرة تسجيل المعلومات بشكل ملحوظ في كل من وادي السند ومصر وبلاد ما بين النهرين (العراق)، وبالتالي زادت دقة المعلومات وأصبح استعمال القياس أمرًا ضروريًّا في الحياة اليومية. وكان لتطور الكتابة في بلاد ما بين النهرين أن أمد البشر بطريقة دقيقة لمتابعة الإنتاج والتعاملات التجارية. إن اللغة المكتوبة مكنت البشر من القياس على أرض الواقع وتسجيله واستعادته في مراحل لاحقة عند الحاجة إليه. إن القياس والتسجيل معًا سهلا من ظهور البيانات، ويُعدان من أقدم أساسيات التحول إلى البيانات.

كان هذا الأمر مهمًّا من أجل تحسين أداء البشر، فقد كانت المباني، على سبيل المثال، من الممكن إنشاؤها طبقًا لسجلات بها أبعادها والمواد المستخدمة في بنائها. سمح هذا الأمر أيضًا بظهور المناهج التجريبية: يمكن لمهندس معماري أو بناء أن يغير من أبعاد بعينها في حين لا يغير أي شيء في أبعاد أخرى، منشئًا بذلك تصميمًا جديدًا ــ والذي بدوره يمكن أن يتم تسجيله. كان بالإمكان أيضًا تسجيل التعاملات التجارية، وبالتالي كان يمكن للمرء أن يعلم كم من المحاصيل قد أنتجت الحقول (وكم من المحاصيل ستأخذه الحكومة كضريبة). مكن القياس من وجود التوقعات وبالتالي التخطيط، حتى وإن كان التخطيط غير ناضج في صورة مجرد تخمين بأن محصول العام القادم سيكون وافرًا كمحصول العام الحالي. سمحت هذه الأنظمة للشركاء بأن يحتفظوا بسجلات عما يدين به كل منهم للآخر. بدون القياس والتسجيل، لم تكن النقود لتظهر؛ لأنه لم تكن لتوجد بيانات لتدعمها.

على مر القرون، توسع القياس من الطول والوزن إلى المساحة والحجم والزمن. مع بداية الألفية الأولى بعد الميلاد، كانت السمات الرئيسية للقياس توجد في الغرب. ولكن كانت هناك عيوب في الطرق التي كانت تستخدمها الحضارات القديمة للقياس، حيث إنها لم تكن تحسن القياس حتى البسيط منه نسبيًّا؛ حيث كان نظام الحساب للترقيم الروماني لا يناسب التحليل الرقمي. بدون نظام ترقيم يعتمد على "مواضع" العشرات أو نظام ترقيم عشري، كان من الصعب أن يتم ضرب أو قسمة الأعداد الكبيرة حتى على الخبراء، وكانت عمليات الجمع والطرح البسيطة تفتقد إلى الوضوح بالنسبة لبقية البشر.

تم تطوير نظام ترقيم بديل في الهند في القرن الأول بعد الميلاد، ومن ثم انتقل إلى بلاد فارس حيث تطور أكثر ، ثم استقر في بلاد العرب الذين حسنوا منه بشكل كبير. وهو أساس نظام الأعداد العربية الذي يستخدمه الغرب في الوقت الحالي. ربما تسببت الحملات الصليبية بدمار الأراضي التي احتلها الأوروبيون، ولكنها ساعدت على هجرة المعرفة من الشرق إلى الغرب، وربما كان من أهمها انتقال نظام الأعداد العربية للغرب. نصح البابا "سلفستر الثاني" الذي درس نظام الأعداد

العربية، باستخدام هذا النظام في نهاية الألفية الأولى بعد الميلاد، وبحلول القرن الثاني عشر تمت ترجمة النصوص العربية التي تصف نظام الترقيم هذا إلى اللاتينية وانتشرت في جميع أنحاء أوروبا؛ مما نتج عنه تطور هائل في الرياضيات في أوروبا.

حتى قبل وصول نظام الأعداد العربية إلى أوروبا، كانت نظام الحساب تطور في أوروبا بفضل استخدام لوحات الحساب، وكانت عبارة عن أطباق توضع عليها عملات رمزية لتدل على الكميات. من خلال تحريك العملات الرمزية من مكان معين لآخر، يمكن للمرء أن يجمع أو يطرح. كان من الصعب حساب الأعداد الكبيرة جدًّا والصغيرة جدًّا في الوقت ذاته. الأمر الأهم أن أرقام هذه اللوحات كانت غير ثابتة، فقد كان من شأن أية حركة خاطئة أو صدمة طائشة أن تغير من الأرقام؛ مما يؤدي إلى نتائج خاطئة. ربما كانت اللوحات الحسابية مقبولة لحساب الأعداد ولكنها كانت سيئة فيما يتعلق بالتسجيل. وكانت الطريقة الوحيدة لتسجيل وتخزين تلك الأرقام هي إعادة تحويلها إلى نظام الترقيم الروماني السيئ. (لم يرى الأوروبيون أبدًا لوحات الحساب الشرقية ــ إن الإدراك المتأخر أمر جيد؛ حيث إن تلك الأجهزة كانت ستساعد على إطالة فترة استخدام نظام الترقيم الروماني في الغرب).

أعطت الرياضيات معنى جديدًا للبيانات؛ حيث أصبح من الممكن الآن **تحليلها**، ليس فقط تسجيلها واستعادتها. استغرق انتشار نظام الأعداد العربية في أوروبا مئات السنين، منذ بداية تقديمها للمجتمع الغربي في القرن الثاني عشر وحتى أواخر القرن السادس عشر. بحلول ذلك الوقت، أعلن علماء الرياضيات أنهم يستطيعون أن يقوموا بالعمليات الحسابية بالأعداد العربية أسرع بست مرات من اللوحات الحسابية. إن ما ساعد الأعداد العربية في النهاية على النجاح في أوروبا هو تطور أداة أخرى للتحول إلى البيانات: نظام قيد الدفاتر المزدوج المدخلات.

اخترع المحاسبون الكتابة في الألفية الثالثة قبل الميلاد، في حين تطور القيد بالدفاتر على مدار القرون التالية لذلك؛ حيث ظلت بشكل عام نظامًا لتسجيل التعاملات التي تحدث في مكان بعينه. الأمر الوحيد الذي فشلت فيه الدفاتر هو أنها لم تكن قادرة في أية لحظة أن تعرض على التجار ما يرغبون في معرفته بسهولة: سواء كان حسابًا بعينه أو رأس مال بأكمله يدر الربح أم لا. بدأ هذا النظام في التغير في القرن الرابع عشر عندما بدأ المحاسبون في إيطاليا بتسجيل التعاملات التجارية باستخدام مدخلين، أحدهما للدائن والآخر للمدين، لذا كانت الحسابات متوازنة بشكل عام. الأمر الرائع في هذا النظام هو أنه جعل من السهل رؤية الأرباح والخسائر. وفجأة بدأت البيانات الساكنة في التعبير عن نفسها.

عادة ما يُعرف نظام القيد بالدفاتر مزدوج المدخلات في الوقت الحالي بترتيبه للحسابات والأمور المالية، ولكنه يعتبر من الأمور المهمة التي ساعدت على انتشار استخدام البيانات؛ حيث مكن من تسجيل البيانات في صورة "فئات" تصل بين الحسابات. لقد عمل هذا النظام بواسطة مجموعة من القواعد عن كيفية تسجيل البيانات ـــ أحد أول الأمثلة الدالة على تسجيل المعلومات القياسي؛ حيث يمكن لأحد المحاسبين أن ينظر في دفاتر محاسب آخر ويفهم ما يوجد به. لقد تم تنظيم هذه الدفاتر لعمل نوع محدد من استيضاح البيانات ـــ حساب الربح والخسارة لكل

حساب ـــ بشكل سريع ومباشر. وفر هذا النظام أيضًا إمكانية مراجعة التعاملات بحيث يكون من السهل تتبع البيانات. قد يقدر المهووسون بالتكنولوجيا هذا الأمر في الوقت الحالي: حيث يحتوي النظام على "تصحيح أخطاء" مدمج كجزء من سمات التصميم. إذا بدا جزء من الدفاتر خاطئًا، أمكن للمرء أن يراجع المدخلات ليكتشف الخطأ.

ولكن، مثله مثل الأعداد العربية، لم يلق قيد الدفاتر مزدوج المدخلات النجاح على الفور، فبعد مائتي عام من ابتكار هذه الطريقة، تطلب الأمر عالم رياضيات وعائلة من التجار ليغيروا من تاريخ التحول إلى البيانات.

كان عالم الرياضيات رجل دين يُدعى "لوكا باتشيولي"، والذي نشر عام 1494 كتابًا تعليميًّا للعامة في الرياضيات وتطبيقاتها التجارية. نجح الكتاب نجاحًا باهرًا وأصبح أفضل كتب الرياضيات في ذلك الوقت. كان هذا الكتاب أيضًا هو الكتاب الأول الذي تُستخدم فيه الأعداد العربية، ومن ثم سهلت شعبية الكتاب من اقتناع الناس بالعمل بالأعداد العربية في أوروبا. كان الأمر الأهم في هذا الكتاب هو القسم الذي تحدث عن القيد في الدفاتر، والذي شرح فيه "باتشيولي" بدقة نظام المدخلات المزدوج في الحسابات. على مدى العقود التالية تم نشر القسم عن قيد الدفاتر وحده بست لغات، وبقي مرجعًا معتمدًا لهذا الموضوع طوال قرون.

أما بالنسبة لعائلة التجار، فقد كانوا أشهر تجار البندقية ورواد الفنون، عائلة "ميديتشي"، والذين أصبحوا في القرن السادس عشر أكثر المصرفيين تأثيرًا في جميع أنحاء أوروبا؛ حيث إنهم استخدموا طريقة رائعة لتسجيل البيانات، نظام المدخلات المزدوجة. قام كل من كتاب "باتشيولي" التعليمي ونجاح عائلة "ميديتشي" في تطبيقه بضمان نجاح قيد الدفاتر مزدوج المدخلات ـــ وبالتالي ثبت من استخدام الأعداد العربية في الغرب.

بالتوازي مع التقدم الحاصل في تسجيل البيانات، كانت طرق قياس العالم ـــ الإشارة للوقت وقياس المسافة والمساحة والحجم والوزن ـــ تصبح أكثر دقة. إن الرغبة الملحة لاستيعاب الطبيعة من خلال القياس هي ما حدد العلم في القرن التاسع عشر، عندما اخترع العلماء أدوات ووحدات جديدة لقياس وتسجيل التيارات الكهربائية والضغط الجوي والحرارة والترددات الصوتية وغيرها. كانت هذه الحقبة هي حقبة إيجاد تعريفات لكل شيء تقريبًا، تحديده وقياسه. تخطى خيال العلماء حدود العقل عندما قاموا بأخذ قياسات لجمجمة الناس كدليل على قدراتهم العقلية. لحسن الحظ اندثرت العلوم الزائفة القائمة على الفراسة، في حين اشتدت الرغبة في القياس أكثر.

تحسن قياس الواقع وتسجيل البيانات بسبب مجموعة الأدوات والعقليات التي تتقبل الأفكار الجديدة. كان هذا المزيج هو التربة الخصبة التي نمت فيها البيانات. كانت مكونات التحول إلى البيانات جاهزة، على الرغم من أنها في العصر التناظري كانت مكلفة وتستغرق الكثير من الوقت. وفي الكثير من الحالات كانت تحتاج إلى صبر لا ينتهي، أو على الأقل إصرار يمتد طوال الحياة؛ مثلما فعل "تيكو براهي" خلال مراقبته المدققة للنجوم والكواكب كل ليلة في القرن السادس عشر. في

الحالات القليلة التي نجح فيها التحول إلى البيانات في الحقبة التناظرية، مثل حالة الضابط البحري "موراي " في الخرائط الملاحية، فإنها عادة ما كانت تنجح بمحض الصدفة ومواتاة الظروف المناسبة: "موراي " على سبيل المثال، تم إسناد عمل مكتبي له مع إمكانية الوصول إلى كنز كبير من السجلات الملاحية, ولكن عندما ينجح التحول إلى البيانات، تنشأ قيمة كبيرة من المعلومات الخفية ويُكشف عن أفكار عظيمة.

تسبب ظهور أجهزة الحاسب في نشوء القياسات الرقمية وأجهزة التخزين التي جعلت التحول إلى البيانات يحدث بشكل أسرع. عملت الحواسب أيضًا على التمكين من التحليل الرياضي للبيانات للكشف عن قيمتها الخفية. باختصار، قامت الرقمنة بدفع التحول إلى البيانات بشكل كبير، ولكنها لا تُعتبر بديلاً لها؛ حيث إن الرقمنة ــ تحويل البيانات التناظرية إلى صيغ يمكن للحواسب قراءتها ــ لا تعتبر تحولًا للبيانات في حد ذاتها.

عندما تتحول الكلمات لتصبح بيانات

إن الاختلاف بين الرقمنة والتحول إلى البيانات صبح جليًّا عندما ننظر إلى نطاق واحد حدث فيه كل منهما ونقارن بين الأمور المترتبة على كل منهما. فكر في الكتب. في عام 2004، أعلنت شركة جوجل عن خطة جريئة للغاية؛ حيث قالت إنها ستأخذ كل صفحة من كل كتاب يمكنها الحصول عليه (في الحدود المسموح بها طبقًا لقانون حماية الملكية الفكرية) وتسمح للجميع من جميع أنحاء العالم أن تبحث في الكتب وتطلع عليها عبر الإنترنت بالمجان. لتحقيق هذه الخطة، تعاونت الشركة مع عدد من كبرى المكتبات الأكاديمية في العالم وطورت أجهزة ماسحات ضوئية من شأنها تحويل الصفحات على الفور إلى الصورة الرقمية، وبذلك كان المسح الضوئي لملايين الكتب ممكنًا وقابلًا للتطبيق من الناحية المادية.

النص الرقمي الأول من جوجل: كانت كل صفحة يتم مسحها ضوئيًّا ووضعها في ملف صور رقمي عالي الجودة والذي كان يُخزن بدوره على خوادم شركة جوجل. تم تحويل الصفحة إلى نسخة رقمية يمكن للناس من جميع أنحاء العالم أن يحصلوا عليها بسهولة من خلال شبكة الإنترنت. كان الحصول على الصفحة يتطلب إما أن يعرف المرء أي الكتب يحتوي على المعلومات التي يرغب في الحصول عليها، أو يقوم بقراءة الكثير من الصفحات حتى يصل إلى المعلومات التي يرغب فيها. لم يكن المرء يستطيع أن يبحث داخل النصوص على كلمات محددة أو يقوم بتحليلها؛ لأن النص لم يكن قد تحول إلى الصورة الرقمية. كل ما فعلته جوجل حينها هو بعض الصور التي يمكن للبشر وحدهم تحويلها إلى معلومات مفيدة ــ من خلال قراءتها.

على الرغم من أن هذه الأداة عظيمة ــ سوف تصبح مكتبة الإسكندرية الحديثة، التي تشتمل على كم هائل من الكتب لم تشمله أكبر المكتبات في التاريخ ــ فإن جوجل أرادت المزيد. أدركت الشركة أن المعلومات قد خزنت قيمًا لن يُطلق لها العنان إلا إذا تم تحويلها إلى بيانات. لذا استخدمت جوجل برنامجًا يعمل بتقنية التعرف البصري على الحروف والذي كان بإمكانه التقاط صور رقمية مع التعرف

على الحروف والكلمات والعبارات داخلها. وكانت النتيجة نصوصًا تم تحويلها إلى بيانات بدلاً من مجرد صور رقمية للصفحات.

أصبحت المعلومات الآن متوفرة ليس فقط للقراء من البشر، ولكن أيضًا لأجهزة الحاسب لمعالجتها والمعادلات الرياضية لتحليلها. مكن التحول للبيانات من جعل النصوص قابلة للفهرسة وبالتالي قابلة للبحث داخلها؛ مما سمح بكمٍّ لا نهائي من تحليل النصوص. يمكننا الآن أن نكتشف ما إذا كانت هناك كلمات أو عبارات معينة استخدمت للمرة الأولى أو أصبحت شائعة. إن المعرفة هي ما يلقي الضوء الجديد على انتشار الأفكار وتطور الفكر الإنساني على مر القرون وبالعديد من اللغات.

يمكنك أن تجرب هذا الأمر بنفسك، برنامج Google Ngram Viewer من شركة جوجل (http://books.google.com/ngram) سيقوم بعمل مخطط بياني عن استخدامات الكلمات أو العبارات طوال الوقت مستخدمًا الفهرس الكامل لموقع كتب جوجل كمصدر للبيانات. في خـــلال ثـــوانٍ أدركنا أنه حتى عام كان مصطلح "السببية" أكثر شيوعًا من مصطلح "العلاقات التبادلية"، ولكن انعكس الأمر بعد ذلك. يمكننا المقارنة بين أساليب الكتابة والحصول على أفكار عن النزاعات بين المؤلفين. سهل التحول إلى البيانات من اكتشاف السرقات الأدبية الأكاديمية، ونتيجة لهذا، أُجبر عدد من السياسيين الأوروبيين، ومن بينهم وزير الدفاع الألماني، على الاستقالة.

يُقدر عدد الكتب الفريدة التي نُشرت منذ اختراع آلة الطباعة في منتصف القرن الخامس عشر بحوالي 130 مليون كتاب. بحلول عام 2012، أي بعد سبع سنوات من بدء مشروع الكتب من شركة جوجل، كانت الشركة قد مسحت ضوئيًّا ما يزيد على 20 مليون كتاب، أي أكثر من 15% من الموروث المكتوب في العالم ــ فرق شاسع. تسبب هذا المشروع في انطلاق نظام أكاديمي جديد أُطلق عليه "Culturomoics": برامج حاسوبية في علم المعاني تحاول استيعاب السلوكيات البشرية وتوجهاتهم الثقافية من خلال التحليل الكمي للنصوص.

في إحدى الدراسات، قام الباحثون من جامعة هارفارد بالتنقيب في ملايين الكتب (أي ما يعادل 500 مليار كلمة) لكشف أن أقل من نصف كلمات اللغة الإنجليزية التي تظهر في الكتب بقليل موجودة بالقواميس. ولكن بدلاً من ذلك، قاموا بكتابة أن وفرة الكلمات "التي تتكون منها المادة السوداء لعلم المعاني لم يتم توثيقها في مراجع مثبتة". علاوة على ذلك، من خلال التحليل الرياضي للمرجعيات الذي قام به الفنان "مارك كاجال"، الذي مُنعت أعماله في ألمانيا النازية, اكتشف الباحثون أن قمع أو مراقبة فكرة ما أو شخص ما يترك "آثارًا قابلة للقياس". إن الكلمات مثل الحفريات المختفية داخل الصفحات بدلاً من الصخور الرسوبية. إن من يمارسون أسلوب Culturomoics يمكنهم أن يستخرجوا تلك الكلمات كما لو كانوا علماء آثار. لا شك في أن المجموعات البيانية تحمل في طياتها أعدادًا لا حصر لها من التوجهات الضمنية ــ هل تعتبر الكتب التي بالمكتبات معبرة عن العالم الحقيقي أو التي يعتز بها المؤلفون وأمناء المكتبات؟ إن أسلوب Culturomoics قد أمدنا بعدسة تمكننا من فهم أنفسنا.

إن تحويل الكلمات إلى بيانات يطلق العنان للكثير من الاستخدامات. لا شك في أن البيانات يمكن استخدامها بواسطة البشر في القراءة وبواسطة الأجهزة في التحليل. ولكن كونها مثالاً يُحتذى به لشركات البيانات الضخمة، تعلم شركة جوجل أن المعلومات لها العديد من الأغراض المحتملة التي يمكن تعديل أسلوب جمعها وتحويلها إلى بيانات؛ لذا استخدمت شركة جوجل النصوص المحولة إلى بيانات بمهارة من مشروع الكتب الممسوحة ضوئيًّا لتحسين نظام ترجمتها الآلية. كما شرحنا في الفصل الثالث، يستخدم النظام الكتب المترجمة من لغات أخرى ويحلل الكلمات والعبارات التي استخدمها المترجمون كبدائل من لغة لأخرى. بإدراك هذا الأمر، تعامل البرنامج مع الترجمة على أنها معادلة حسابية ضخمة، مع استيعاب الحاسب الآلي للاحتمالات ليتمكن من تحديد أي الكلمات تعد أفضل بديل لكلمات أخرى عند الترجمة من لغة لأخرى.

لا شك في أنه لم تكن شركة جوجل وحدها هي التي رغبت في إحضار المحتوى الغني للموروث المكتوب في العالم إلى عالم الحواسب الآلية، وكانت هي الشركة الأولى التي تحاول فعل ذلك الأمر بفارق بسيط عن مشروع آخر. مشروع "جوتنبرج"، مبادرة تطوعية لوضع نطاق عام يعمل عبر الإنترنت منذ عام 1971، وكان يدور حول جعل النصوص متوفرة للأشخاص عبر الإنترنت لقراءتها، ولكنه لم يفكر في الاستخدامات الأخرى للتعامل مع النصوص على أنها بيانات. كان المشروع يدور حول القراءة لا إعادة الاستخدام. بالمثل، حاول الناشرون على مدى أعوام طوال أن ينتجوا نسخًا إلكترونية من الكتب؛ حيث أدركوا هم أيضًا قيمة الكتب من حيث المحتوى، لا كبيانات ــ وكان نموذج عملهم يقوم على هذا الأمر. وبالتالي لم يستخدموا أبدًا أو يسمحوا للآخرين باستخدام الموروث البياني المختفي داخل نصوص الكتب. لم يروا أبدًا الحاجة أو يقدروا الاحتمالات.

تتنافس شركات كثيرة في الوقت الحالي للسيطرة على سوق الكتب الإلكترونية. شركة أمازون، مع قرائها المتحمسين للكتب الإلكترونية بصيغة Kindle، يبدو أنها في موقع الصدارة بفارق كبير. ولكن هذا هو المجال الذي تختلف فيه إستراتيجيات كل من أمازون وجوجل بشكل كبير.

قامت شركة أمازون هي الأخرى بتحويل الكتب إلى بيانات ــ ولكن على العكس من جوجل، فشلت في إيجاد استخدامات جديدة للنصوص بعد تحويلها إلى بيانات. أقنع "جيف بيزوس"، مؤسس الشركة ومديرها التنفيذي، مئات الناشرين بأن ينشروا كتبهم بصيغة Kindle. كتب Kindle ليست عبارة عن صور لصفحات الكتاب؛ حيث إنها لو كانت صورًا لم يكن المرء قادرًا على تغيير حجم الخط أو عرض الصفحات ملونة بالإضافة إلى الأبيض والأسود. إن النص قد تم تحويله إلى بيانات، ولم تتم رقمنته فحسب. لا شك في أن شركة أمازون قد قامت مع الكتب الجديدة بما سعت شركة جوجل حثيثًا للقيام به مع الكثير من الكتب القديمة.

ولكن، وبعيدًا عن خدمات أمازون العبقرية الأخرى المتعلقة بـ "الكلمات المهمة إحصائيًّا" ــ التي تستخدم المعادلات الرياضية لإيجاد صلات بين موضوعات الكتب التي قد لا تكون واضحة ــ لم يستخدم متجر بيع الكتب عبر الإنترنت ثروة الكلمات

تلك في تحليل البيانات الضخمة؛ حيث إن الشركة ترى عملها في مجال بيع الكتب مبنيًّا على المحتوى الذي يقرؤه البشر، بدلاً من تحليل الحواسب لنصوص محولة بيانيًّا. ومن الإنصاف أن نقول إن الشركة ربما واجهت قيودًا من ناشرين محافظين على الكيفية التي ستستخدم الشركة بها المعلومات المذكورة في كتبهم. أما شركة جوجل، الفتى الشقي في عالم البيانات الضخمة، فتميل دومًا إلى تخطي جميع الحدود، لا تشعر بوجود أية قيود عليها: إنها تحصل على قوت يومها من ضغطات المستخدمين، ليس من خلال الدخول إلى كتب الناشرين. ربما يكون من المنصف أن نقول، على الأقل في الوقت الحالي، إن أمازون تدرك قيمة المحتوى الرقمي، في حين تدرك جوجل قيمة تحويله إلى بيانات.

عندما يتحول الموقع إلى بيانات

من بين أهم المعلومات الموجودة في العالم، العالم نفسه. ولكن على الرغم من ذلك فإنه في أغلب حقب التاريخ لم يتم استخدام الحيز المكاني كبيانات. إن الموقع الجغرافي للطبيعة والأشياء والأشخاص يحتوي بلا شك على الكثير من المعلومات. الجبل يقع هنا والشخص يقف هنا. ولكن الأمر الأكثر أهمية هو أن المعلومات تحتاج إلى تحويلها إلى بيانات. يتطلب الأمر لتحويل الموقع إلى بيانات بعض الاستعدادات المسبقة؛ حيث إننا بحاجة إلى طريقة لقياس كل بوصة مربعة من مساحة الكرة الأرضية. ونحتاج إلى طريقة معيارية لتدوين القياسات. ونحتاج إلى أداة لمتابعة وتسجيل البيانات. القياس والمعايرة والتجميع. عند توفير جميع هذه الاحتياجات فقط يمكننا أن نخزن ونحلل المواقع ليس كمكان، بل كبيانات.

في الغرب، بدأ قياس المكان في اليونان القديمة؛ ففي حوالي عام 200 قبل الميلاد، اخترع "إيراتوستينيس" نظامًا من المربعات لتحديد الأماكن بناءً على العرض والطول. ولكن مثلما يحدث مع الكثير من الأفكار الجيدة القديمة، اندثرت هذه الفكرة بمرور الوقت. وبعد ألف وخمسمائة عام، في حوالي عام 1400 ميلادية، وصلت نسخة من كتاب "بطليموس" تحت عنوان "**الجغرافيا**" إلى فلورنسا من القسطنطينية عندما أشعلت النهضة الأدبية والتجارة البحرية فتيل الاهتمام بالعلوم والاستعانة بالمعرفة من الأقدمين. كانت دراسة "بطليموس" رائعة، وتم تطبيق دروسه القديمة لحل مشكلات الملاحة الحديثة. من هذا الوقت فصاعدًا، ظهرت الخرائط بخطوط الطول والعرض ومقياس للخريطة. تم تطوير هذا النظام أكثر فيما بعد بواسطة رسام الخرائط الفلمنكي "جيراردوس ميركاتور" عام 1570، والذي مكن البحارة من الإبحار في خط مستقيم في العالم كروي الشكل.

على الرغم من وجود بعض الوسائل لتسجيل المعلومات في هذا الوقت، فإنه لم توجد أية صيغ مقبولة بوجه عام لمشاركة المعلومات؛ فقد كنا بحاجة إلى نظام مطابقة عام، مثلما استفادت شبكة الإنترنت من أسماء النطاقات لجعل أشياء مثل البريد الإلكتروني تعمل في جميع أنحاء العالم. استغرق جعل خطوط الطول والعرض معيارًا وقتًا طويلاً؛ حيث تم الاعتراف بها في عام 1884 خلال مؤتمر خطوط الطول الدولي في العاصمة الأمريكية واشنطن حيث اختارت 25 دولة مدينة

جرينيتش بإنجلترا، كخط الطول الرئيسي ونقطة الصفر (مع امتناع الفرنسيين الذين كانوا يعتبرون أنفسهم قادة المعايير الدولية). في أربعينيات القرن العشرين، تم ابتكار نظام الخرائط المستعرضة للكرة الأرضية الذي قسم الأرض إلى 60 منطقة لزيادة الدقة.

يمكن الآن تحديد الموقع الجغرافي المكاني وتخزينه وحسابه وتحليله وتوصيله إلى الآخرين في صيغة معيارية عددية. يمكن تحويل المواقع إلى بيانات، ولكن بسبب التكلفة العالية لقياس وتسجيل المعلومات في الصيغ التناظرية، كان هذا الأمر نادر الحدوث. للتمكن من التحول إلى البيانات للمواقع، كان يجب اختراع الأدوات التي يمكنها قياس المواقع بتكلفة زهيدة. حتى سبعينيات القرن العشرين كانت الطريقة الوحيدة لتحديد المواقع على الطبيعة هي نقاط الاستدلال والاسترشاد بالنجوم وتحديد مواقع السفن وتكنولوجيا محدودة لتحديد المواقع بموجات الراديو.

حدث تحول كبير عام 1978؛ حيث تم إطلاق القمر الصناعي الأول من بين الأقمار الأربعة والعشرين التي تشكل نظام تحديد المواقع العالمي. يمكن للمستقبلات الأرضية تثليث مكانها من خلال تحديد الاختلافات في الوقت التي يستغرقها استقبال إشارة من قمر صناعي يبعد 12600 ميل فوق رءوسنا. النظام الذي تم تطويره بواسطة وزارة الدفاع الأمريكية، تم استخدامه للمرة الأولى في ثمانينيات القرن العشرين لأغراض غير عسكرية وأصبح يعمل بكامل طاقته في تسعينيات القرن العشرين. تم تحسين دقة النظام من أجل التطبيقات التجارية بعد عقد كامل، ومع وصول دقة النظام إلى متر واحد، أصبح نظام تحديد المواقع العالمي هو اللحظة الفاصلة عندما ظهرت طريقة لتحديد المواقع، حلم جميع الملاحين وراسمي الخرائط والرياضيين منذ الأزل، بواسطة الوسائل التكنولوجية لتحديدها بسرعة كبيرة ورخيصة (نسبيًّا)، ولا تحتاج إلى أية معرفة متخصصة.

على الرغم من أن المعلومات يجب إنتاجها، إلا أنه لم يكن هناك ما يمنع كلاًّ من "إيراتوستينيس" و"ميركاتور" من تقدير مواقعهم في كل دقيقة من اليوم، إذا رغبوا في ذلك. على الرغم من إمكانية تطبيق هذا الأمر إلا أنه لم يكن عمليًّا. وبالمثل، كانت أجهزة استقبال نظام تحديد المواقع العالمي معقدة ومكلفة، وكانت ملائمة للغواصات ــ ليس للجميع طوال الوقت. ولكن هذا الأمر من الممكن تغييره بفضل وفرة الرقاقات الإلكترونية رخيصة الثمن التي تم إدراجها في الأجهزة الرقمية. انخفضت تكلفة وحدة نظام تحديد المواقع العالمي من مئات الدولارات في التسعينيات إلى دولار واحد في الوقت الحالي على أقصى تقدير. عادة ما يستغرق نظام تحديد المواقع العالمي بضع ثوان لتحديد موقع ما، وأصبحت الإحداثيات معيارًا للعمل؛ لذا قد تعني الإحداثيات التالية: 37 درجة 14 دقيقة 6 ثوان شمالاً، 115 درجة 48 دقيقة 40 ثانية غربًا، أن المرء يوجد في واحدة من أكثر القواعد العسكرية الأمريكية سرية في بقعة منعزلة من صحراء نيفادا يُطلق عليها "المنطقة 51"؛ حيث (ربما) يحتفظون بالمخلوقات الفضائية.

يعتبر نظام تحديد المواقع العالمي واحدًا من بين أنظمة كثيرة لتحديد المواقع؛

حيث هناك عدد من أنظمة الأقمار الصناعية يتم تجهيزها في الصين وأوروبا. ويمكن أن تتحسن الدقة من خلال الربط بين أبراج شبكات الهواتف المحمولة أو موزعات الشبكات اللاسلكية لتحديد الموقع بناءً على قوة الإشارة؛ حيث إن نظام تحديد المواقع العالمي لا يعمل داخل البيوت أو في منتصف البنايات العالية. يساعد هذا الأمر على التحقق من سبب قيام شركات مثل جوجل وأبل ومايكروسوفت بإنشاء نظام تحديد المواقع جغرافيًّا خاصًّا بكل منها ليكمل نظام تحديد المواقع العالمي؛ فبرنامج Street View من شركة جوجل للسيارات يجمع معلومات من موزعات الشبكات اللاسلكية عند التقاطها للصور، وجهاز الآي فون به برنامج "جاسوس الهاتف" الذي يجمع معلومات عن الموقع وبيانات الشبكة اللاسلكية ويرسلها مرة أخرى إلى شركة أبل دون أن يدرك المستخدمون ذلك. (نظام تشغيل أندرويد من جوجل ونظام تشغيل الهواتف المحمولة من شركة مايكروسوفت يجمعان أيضًا هذه البيانات).

الآن، لم يعد الأشخاص فقط هم من يمكن تتبعهم، بل الأشياء أيضًا. من خلال الوحدات الإلكترونية الموضوعة داخل المركبات، فإن التحول البياني للمواقع من شأنه أن يحول فكرة التأمين بأكملها؛ حيث تقدم البيانات نظرة متعمقة للأوقات والمواقع والمسافات للقيادة الحقيقية للتقليل من هامش المخاطرة في تحديد الأسعار. في الولايات المتحدة الأمريكية وبريطانيا، يمكن للسائقين أن يشتروا التأمين على السيارات بسعر يُحدد وفقًا للمكان والوقت الذي يقودون فيهما سيارتهما بالفعل، أي إنهم لن يدفعوا نسبة سنوية بناءً على أعمارهم ونوعهم وسجلهم السابق. إن أساليب تحديد أسعار التأمين تلك تشكل حافزًا للسلوكيات الجيدة؛ حيث إنها تحول طبيعة التأمين من الاعتماد على مجموع المخاطر إلى طبيعة أخرى تعتمد على التصرف الفردي لكل شخص. إن مراقبة الأشخاص من خلال المركبات يعمل أيضًا على تغيير التكاليف الثابتة، مثل الطرق وغيرها من البنى التحتية، من خلال ربط استخدام هذه الموارد بالسائقين وغيرهم ممن "يستهلكونها". كان هذا الأمر من المستحيل تحقيقه قبل توفير نظام تحديد المواقع الجغرافية في الشكل البياني بصورة مستمرة عن جميع الأشخاص والأماكن ـــ ولكن هذا هو العالم الذي نتوجه إليه.

شركة يو بي إس للشحن على سبيل المثال، تستخدم بيانات "المواقع الجغرافية" بطرق عدة؛ حيث إن مركباتها مزودة بحساسات ووحدات لاسلكية ونظام تحديد المواقع العالمي مما يسمح للمقر الرئيسي بأن يتوقع مشكلات المحركات، كما رأينا في الفصل السابق. علاوة على ذلك، يمكن للمقر الرئيسي تحديد مكان وجود الشاحنات في حالة تأخرها، ومراقبة الموظفين ومتابعة مسار الرحلات لتحديد المسارات. يتم تحديد مسار الرحلة المثالي كجزء من بيانات رحلات توصيل سابقة، بما يشبه إلى حد كبير خرائط "موراي " التي كانت تعتمد على الرحلات البحرية القديمة.

كان لبرنامج التحليل تأثير كبير؛ ففي عام 2011، غطى سائقو شركة يو بي إس حوالي 30 مليون ميل من الطرق، مع توفير 3 ملايين جالون من الوقود و30 ألف

طن متري من انبعاث ثاني آكسيد الكربون، طبقًا لما قاله "جاك ليفيس" مدير عمليات الشركة. طور البرنامج أيضًا من الأمان والكفاءة: كانت المعادلات تقدم الطرق ذات الدورانات الأقل التي يجب أن تمر بها السيارات عند التقاطعات، والتي قد تؤدي إلى وقوع حوادث وإضاعة الوقت واستهلاك المزيد من الوقود حيث إن الشاحنات يجب أن تكون على وضع عدم الحركة عند الدوران.

يقول "ليفيس" في شركة يو بي إس: "إن التوقعات قد أمدتنا بالمعرفة، ولكن بعد المعرفة هناك أمر إضافي: الحكمة والبصيرة. في مرحلة ما من الزمن، سيصبح النظام ذكيًّا ومتطورًا لدرجة أنه سيتوقع المشكلات ويصححها قبل أن يدرك المستخدم وجود المشكلة من الأساس".

إن المواقع في شكلها البياني على مدار الزمن قد تم تطبيقها بشكل ملحوظ على البشر. طوال سنوات كانت شركات الشبكات اللاسلكية تقوم بجمع وتحليل المعلومات لتحسين مستوى خدمات شبكاتها. ولكن تزايد استخدام البيانات في أغراض أخرى وأصبح يتم جمعها من أطراف أخرى للخروج بخدمات جديدة. بعض تطبيقات الهواتف الذكية، على سبيل المثال، تقوم بجمع معلومات المواقع بغض النظر عما إذا كان التطبيق ذاته يحتوي على سمة تحديد المواقع أم لا. في حالات أخرى، يكون الغرض من التطبيق بأكمله هو إنشاء عمل كامل حول مكان تواجد المستخدم، مثال على ذلك، تطبيق فورسكوير الذي يمكن المستخدمين من "الحجز" في الأماكن التي يفضلونها، ويحصل هذا البرنامج على أجر من برامج الولاء والتوصية بالمطاعم وغيرها من الخدمات المتعلقة بالمواقع.

إن القدرة على جمع بيانات الموقع الجغرافي للمستخدمين أصبحت مهمة للغاية؛ حيث إنها، على المستوى الفردي، تسمح بوجود الدعاية الهادفة التي تعتمد على مكان وجود الشخص أو المكان الذي من المتوقع الذهاب إليه. هذا بالإضافة إلى أن المعلومات يمكن تكديسها لتكشف عن توجهات المستخدمين. على سبيل المثال، يمكن لبيانات المواقع المكدسة أن تسمح للشركات بتوقع حدوث الاختناقات المرورية دون رؤية السيارات نفسها: عدد وسرعة الهواتف التي تتحرك على الطريق السريع من شأنها الكشف عن تلك المعلومات. تقوم شركة "إير سايج" بتحليل 15 مليار سجل من بيانات المواقع الجغرافية يوميًّا من تنقلات الهواتف المحمولة لملايين المشتركين من أجل إنشاء تقارير مرورية فورية لحوالي 100 مدينة في الولايات المتحدة الأمريكية. هناك شركتان أخريان هما "سنس نتووركس" و"سكا يهوك"، اللتان تستخدمان بيانات المواقع الجغرافية لتحديد أي الأماكن في المدينة تعتبر أفضل الأماكن لقضاء السهرات، أو لتقدير عدد المحتجين المتواجدين في إحدى المظاهرات.

ولكن، تبين أن الاستخدامات غير التجارية لبيانات المواقع الجغرافية هي أهم الاستخدامات. قامت "ساندي بنتلاند"، مديرة معمل التحركات البشرية بمعهد ماساتشوستس للتقنية، و"ناثان إيجل" بابتكار ما يطلقان عليه "التنقيب الواقعي". الأمر الذي يشير إلى معالجة كم هائل من البيانات من الهواتف المحمولة للتوصل إلى استنتاجات وتوقعات عن السلوك البشري. في إحدى الدراسات، سمح تحليل

التحركات وأنماط المحادثات الهاتفية لهم بأن يحددوا بنجاح الأشخاص الذين أصيبوا بعدوى الأنفلونزا قبل حتى أن يعلموا هم أنفسهم ذلك. في حالة وباء الأنفلونزا المميت، يمكن لهذه القدرة أن تحافظ على حياة الملايين من خلال جعل مسئولي الصحة العامة يدركون أكثر الأماكن المنكوبة في أية لحظة. ولكن إذا وصلت هذه القدرة إلى أيدي غير مسئولة، فيمكن لقوة التنقيب الواقعي أن يكون لها عواقب وخيمة كما سنرى لاحقًا.

قام "إيجل"، مؤسس شركة بيانات لاسلكية تحت اسم "جانا"، باستخدام بيانات هواتف محمولة مكدسة بما يزيد على 200 شركة شبكات هواتف محمولة مما يزيد على 100 دولة ــ حوالي 3.5 مليار إنسان من أمريكا اللاتينية وإفريقيا وأوروبا ــ للإجابة عن أسئلة كان يتمنى مديرو التسويق الحصول على إجابة عنها، مثل كم عدد المرات التي تقوم ربة المنزل بغسيل الملابس كل أسبوع. ولكنه استخدم أيضًا هذه البيانات الضخمة من أجل البحث عن إجابة لأسئلة على غرار كيف تزدهر المدن؛ حيث قام هو وزميل له بدمج بيانات المواقع من مشتركي الهواتف المحمولة المدفوعة مسبقًا في إفريقيا مع كم الأموال التي دفعوها عندما أوقفوا اشتراكاتهم. كانت هناك علاقة تبادلية قوية بين القيمة والدخل: الأشخاص الأغنى يشترون دقائق أكثر في كل مرة. ولكن من بين اكتشافات "إيجل" المنافية للمنطق أن الأحياء الفقيرة، بدلًا من أن تكون مراكز لتجمع الفقراء، كانت تعمل أيضًا كنقطة للانطلاق الاقتصادي. الفكرة هي أن هذه الاستخدامات غير المباشرة لبيانات المواقع لا توجد صلة بينها وبين مسارات الاتصالات الهاتفية الخلوية، الغرض الذي تم جمع المعلومات من أجله في المقام الأول. ولكن بدلاً من هذا، بمجرد أن يتم تحويل المواقع إلى بيانات، يمكن أن تُبتكر استخدامات جديدة ويمكن أن تنشأ قيم جديدة.

عندما يتحول التعامل بين الناس إلى بيانات

إن المجالات التالية التي ستتحول إلى بيانات ستكون شخصية أكثر: علاقاتنا وخبراتنا ومزاجنا. إن فكرة التحول إلى بيانات هي حجر الأساس الذي تقوم عليه الكثير من شركات مواقع التواصل الاجتماعي. إن منصات شبكات التواصل الاجتماعي لا توفر لنا طرقًا لإيجاد الأصدقاء والزملاء والحفاظ على التواصل معهم فحسب، بل تتناول العناصر غير الملموسة من حياتنا اليومية وتحولها إلى بيانات لتسمح باستخدامها في القيام بأمور جديدة. قام موقع فيس بوك بتحويل العلاقات البشرية إلى بيانات؛ التي كانت متواجدة منذ الأزل وتحتوي على معلومات، ولكنها لم تُعرف على أنها بيانات حتى ظهور "الرسم البياني الاجتماعي" لموقع فيس بوك. مكن موقع تويتر من تحويل المشاعر إلى بيانات من خلال إنشاء طريقة تمكن الناس من تسجيل ومشاركة أفكارهم الشاردة، والتي كانت فيما سبق تُفقد بمرور الوقت. حول موقع "لينكد إن "خبراتنا الطويلة السابقة إلى بيانات، مثلما حول "موراي " سجلات السفن القديمة إلى بيانات، وحُوِّلت هذه المعلومات إلى توقعات عن حاضرنا ومستقبلنا: التي قد نعلمها، أو وظائف قد نرغب في الالتحاق بها.

ما زالت استخدامات البيانات تلك في طور النمو. في حالة موقع فيس بوك، كانت الشركة صبورة وذات نظرة ثاقبة إلى حد كبير؛ حيث إنها كانت تعلم أن الكشف عن الكثير من أغراض البيانات للمستخدمين في وقت مبكر سيجعلهم يشعرون بالارتباك. هذا إلى جانب أن الشركة ما زالت تعدل من نموذج عملها (وسياسة الخصوصية) فيما يتعلق بكمية ونوع البيانات التي ترغب في جمعها. لذا فإن أغلب النقد الذي وجه إلى الشركة كان يركز على نوع المعلومات التي يمكنها جمعه بدلاً من الاستخدامات الفعلية التي وظفت فيها هذه البيانات. كان عدد مستخدمي فيس بوك عام 2012 مليار مستخدم، الذين كونوا علاقات صداقة بلغ عددها 100 مليار علاقة صداقة. أظهر الرسم البياني الاجتماعي أن ما يزيد على عُشر إجمالي عدد سكان العالم تم تحويل علاقاتهم إلى بيانات وأن هذه البيانات متوفرة لشركة واحدة فقط.

إن الاستخدامات المحتملة لهذه البيانات لا حد لها. فكرت بعض الشركات الجديدة في استخدام الرسم البياني الاجتماعي كمؤشر لإنشاء تقارير ائتمانية. الفكرة تقوم على أن الطيور على أشكالها تقع: الأشخاص الحريصون يصادقون أشخاصًا ذوي أفكار مماثلة، في حين يتعرف المستهترون على من يماثلهم. يمكننا أن نقول إن موقع فيس بوك نجح في أن يكون بمثابة شركة "فيكو"، شركة التقارير الائتمانية الجديدة. إن الكم الكبير للمجموعات البيانية من شركات التواصل الاجتماعي من شأنها أن تشكل أساسًا جيدًا لشركات جديدة تتخطى الاستخدام السطحي القائم على مشاركة الصور وتحديث الحالات و"الإعجاب".

موقع تويتر أيضًا شهد استخدام بياناته بطرق مشوقة. بالنسبة للبعض، أن يتم إرسال 400 مليون تغريدة مختصرة كل يوم عام 2012 بواسطة 140 مليون مستخدم شهري أمر لا يمكن أن يكون تافهًا أو عشوائيًّا. وفي الحقيقة، عادة ما يكون كذلك. ولكن قامت الشركة بتحويل أفكار ومزاج وتعاملات الناس إلى بيانات، الأمر الذي لم يكن ممكنًا فيما سبق. أبرمت شركة تويتر عقدين مع شركتين هما "داتا سيفت" و"جنيب"، من أجل منحهم حق الحصول على بياناتها (على الرغم من أن جميع التغريدات علنية، فإن الولوج للمنبع له ثمن). تقوم الكثير من الشركات بتحليل التغريدات، أحيانًا بواسطة أسلوب يُطلق عليه تحليل المشاعر؛ لتخزين معلومات مكدسة عن العملاء أو الحكم على فعالية حملات التسويق.

هناك اثنتان من شركات المحافظ الوقائية، "ديروينت كابيتال" في لندن و"ماركت سايك" في كاليفورنيا، بدأتا في تحليل نصوص التغريدات المحولة إلى بيانات كمؤشرات للاستثمار في سوق الأوراق المالية (تم الاحتفاظ بالإستراتيجيات التجارية الفعلية الخاصة بهما سرًّا: بدلاً من الاستثمار في الشركات التي تقوم بدعاية صاخبة، يقومون بوضع عروض مضادة لها). تقوم كلتا الشركتين في الوقت الحالي ببيع المعلومات للمضاربين. في حالة شركة "ماركت سايك"، تعاونت مع شركة "تومسون رويترز" لتقديم ما لا يقل عن 18.864 مؤشر في 119 دولة، ويتم تحديثها كل دقيقة، بناءً على الحالات العاطفية مثل التفاؤل والعبوس والفرحة والخوف والغضب وحتى الأفكار الرئيسية مثل الابتكار والخصومة والنزاع. لا يتم

استخدام هذه البيانات بشكل كبير من قبل البشر بل من قبل الحواسب: عباقرة الرياضيات في شركات وول ستريت، الذين يُطلق عليهم "كوانتس"، يغذون نماذجهم الخوارزمية بالبيانات ليكتشفوا العلاقات التبادلية الخفية التي يمكن استثمارها للتحول إلى أرباح. إن تكرار التغريدات عن موضوع معين من شأنه أن ينبئ بالكثير، مثل عائد أفلام هوليود، طبقًا لما قاله أحد رواد تحليل مواقع التواصل الاجتماعي "برناندو هوبرمان"؛ حيث قام وأحد زملائه من شركة "إتش بي" بتطوير نموذج يتابع معدل نشر التغريدات، ومن خلال هذا تمكنا من توقع نجاح الأفلام أفضل من وسائل التوقع التقليدية المستخدمة.

ولكن هناك المزيد من الاحتمالات. إن رسائل تويتر مقيدة بعدد محدود من الحروف يبلغ 140 حرفًا، ولكن الميتاداتا ـــ أي "المعلومات عن المعلومات" ـــ المصاحبة لكل تغريدة تكون غنية للغاية. قد لا يبدو بعضها مفيدًا بدرجة كبيرة، مثل "الواجهة الرئيسية" التي تظهر على صفحة المستخدم على تويتر أو على البرنامج الذي يدخل من خلاله المستخدم إلى الخدمة. ولكن هناك ميتاداتا أخرى تكون مشوقة للغاية، مثل لغة المستخدم، وموقعه الجغرافي، وعدد وأسماء الأشخاص الذين يتبعونه على الموقع وأيضًا الذين يتبعهم. في إحدى الدراسات العلمية المنشورة في جريدة "**العلم**" عام 2011، أظهر تحليل 509 مليون تغريدة على مدار عامين نشرها 2.4 مليون شخص من 84 دولة أن مزاج الناس يتبع أنماطًا يومية أو أسبوعية في مختلف الثقافات حول العالم ـــ الأمر الذي كان من المستحيل ملاحظته من قبل. لقد تم تحويل مزاج الناس إلى بيانات.

إن التحول إلى البيانات لا يدور حول تحويل السلوكيات والمشاعر إلى صورة يمكن تحليلها فحسب، بل أيضًا السلوك البشري ككل. هذا الأمر من الصعب متابعته، خاصة في سياق المجتمع الأكبر والمجموعات الفرعية التي يحتوي عليها. قام كل من عالم الأحياء "مارسل سالاذي" من جامعة ولاية بنسلفانيا ومهندس البرمجيات "شاشانك خاندلوال" بتحليل التغريدات ليجدوا أن سلوكيات الناس حول الإجازات تتفق مع احتمالية تلقيهم لحقن الأنفلونزا. الأمر المهم، هو أنهم استخدموا الميتاداتا المتعلقة بمن يتصل مع من في شبكة متابعي تويتر ليقوموا بأمر آخر. لاحظوا أن احتمالية وجود مجموعة فرعية من الأشخاص الذين لم يتلقوا اللقاح واردة. إن ما يجعل هذه الدراسة مميزة هو أنه في حين استخدمت الدراسات الأخرى، مثل "فلو ترندز" من شركة جوجل، البيانات المكدسة لمعرفة الحالة الصحية للأفراد، قام تحليل المشاعر الذي أجراه "سلاذي" بتوقع **السلوكيات** الصحية بشكل صحيح.

هذه الاكتشافات المبكرة تشير إلى أين سيصل التحويل البياني في المستقبل دون أدنى شك. على غرار جوجل، هناك مجموعة من مواقع التواصل الاجتماعي مثل فيس بوك وتويتر ولينكد إن وفور سكواير وغيرها ممن تجلس على كنز من المعلومات المحولة إلى بيانات، ستلقي الضوء على التحركات الاجتماعية بجميع مستوياتها، بدءًا من الفرد وصولاً إلى المجتمع ككل.

تحويل كل شيء إلى بيانات

بالقليل من الخيال، هناك عدد كبير من الأشياء التي يمكن تحويلها إلى بيانات ــ والتي تفاجئنا على طول الطريق. بنفس الطريقة التي عمل بها البروفيسور "كوشيميزو" في طوكيو على المقاعد، حصلت شركة "آي بي إم" على حق اختراع في الولايات المتحدة الأمريكية عام 2012 عن "تأمين المباني باستخدام تكنولوجيا الحواسب القائمة على المساحة". يدور الاختراع حول أرضية كاملة لأدوار المباني حساسة للمس، مثل شاشة هاتف ذكي عملاقة. كانت الاستخدامات المحتملة لهذا الاختراع كثيرة؛ حيث إنه قادر على تحديد الأشياء التي تلامسه. في الأساس، يمكنه أن يضيء الأنوار في الغرفة أو يفتح الباب عند قدوم شخص ما. الأمر الأهم، هو أنه قادر على تحديد الأشخاص من أوزانهم أو طريقة وقوفهم وسيرهم. يمكنه أن يستدعي الطوارئ إذا ما سقط شخص ما على الأرض ولم ينهض، وهي خاصية مهمة لكبار السن. يمكن أصحاب المحلات أن يعرفوا كيفية سير حركة البيع والشراء في محلاتهم. عندما يتم تحويل الأرضيات إلى بيانات، لن يكون هناك سقف لاستخداماتها المحتملة.

إن احتمالية تحويل كل شيء إلى بيانات ليست صعبة كما يبدو. تذكر حركة "الذات المقاسة"، التي تشير إلى مجموعة متفاوتة من المهووسين باللياقة البدنية والحالة الصحية والمدمنين على التكنولوجيا الذين يقيسون كلًّا من أعضاء أجسامهم ويحيون من أجل تحسين الحياة ــ أو على الأقل من أجل تعلم أمور جديدة لم يكونوا ليعرفوها من قبل. إن عدد "المراقبين لأنفسهم" قليل حاليًّا ولكنه آخذ في التزايد.

بفضل الهواتف الذكية وتقنيات الحوسبة الرخيصة، لم يكن من الممكن تحويل أكثر السلوكيات الحياتية أساسية إلى بيانات أسهل من ذلك؛ حيث سمح عدد كبير من الشركات الناشئة للناس بأن يتتبعوا أنماط نومهم من خلال قياس ذبذبات عقولهم في أثناء النوم. قامت إحدى الشركات، زيو، بإنشاء أكبر قاعدة بيانات في العالم عن أنشطة النوم وأماطت اللثام عن اختلاف حركات العينين السريعة في أثناء النوم بين الرجال والنساء. قامت شركة "اسثيمابوليس" بتركيب حساسات إلى أجهزة استنشاق مرضى الربو والتي تحدد المواقع عن طريق نظام تحديد المواقع العالمي؛ حيث ساعدت المعلومات المتراكمة الشركة على تحديد المؤثرات البيئية التي تسبب نوبات الربو، مثل الاقتراب من نباتات معينة.

ساعدت شركتا "فيتبيت" و"جاوبون" الناس على قياس نشاطهم البدني في أثناء النوم. هناك شركة أخرى وهي "بايسس" جعلت من يرتدون أربطة المعصم التي تنتجها يتابعون علاماتهم الحيوية، بما فيها معدل ضربات القلب وتغير قدرة بشرتهم على توصيل الكهرباء، الذي يعد من علامات التوتر. لقد أصبح الحصول على البيانات أكثر سهولة وتدخلاً في جميع المجالات مما مضى. في عام 2009، حصلت شركة أبل على براءة اختراع عن جمع البيانات عن نسبة الأكسجين بالدم ومعدل ضربات القلب ودرجة حرارة الجسم من خلال سماعات الأذنين التي تنتجها.

هناك الكثير مما يمكننا معرفته من تحويل كيفية عمل أجسامنا إلى بيانات. قام الباحثون بجامعة "جوفيك" بالنرويج بالتعاون مع شركة "ديراوي" للإحصاءات

الحيوية بتطوير تطبيق للهواتف الذكية يحلل طريقة سير المستخدم ويستخدم هذه المعلومات كبرنامج أمان يغلق الهاتف عند سرقته. في الوقت ذاته، قام أستاذان من معهد جورجيا للأبحاث التقنية هما "روبرت ديلانو" و"بريان باريس" بتطوير تطبيق للهواتف الذكية يدعى "آي تريم" والذي يستخدم مقياس السرعة المدرج في الهاتف لمتابعة الرعشات التي تصيب الجسم بسبب متلازمة باركنسون وغيرها من أمراض التهاب الأعصاب. كان هذا التطبيق نعمة لكل من الأطباء والمرضى؛ حيث إنه مكن المرضى من تجنب الفحوصات الباهظة التي كانت تجرى في عيادات الأطباء، ومكن الأطباء من متابعة عجز المرضى واستجابتهم لطرق العلاج عن بعد. طبقًا للباحثين في كيوتو، لا تقل فعالية الهواتف الذكية كثيرًا في قياس الرعشات عن جهاز قياس السرعة ثلاثي المحاور المستخدم في العيادات المتخصصة؛ لذا يمكن الاعتماد عليه. مرة أخرى أقول: إن القليل من الفوضى تزيد من الدقة.

في أغلب تلك الحالات، نقوم بجمع المعلومات وتحويلها إلى بيانات حتى يمكن إعادة استخدامها. يمكن أن يحدث هذا في كل مكان تقريبًا وينطبق على كل شيء تقريبًا. تبيع شركة "جرين جوز"، شركة حديثة العهد في سان فرانسيسكو، الحساسات الدقيقة التي تتبع الحركة، والتي يمكن تركيبها على أشياء لتتبع كم يتم استخدامها. ويمكن وضعها في علبة خيوط تنظيف الأسنان أو صندوق فضلات القطط أن يحول العناية بالأسنان والنباتات والحيوانات الأليفة إلى بيانات. إن الحماسة الموجهة "لشبكة الأشياء" ــ الرقاقات الإلكترونية والحساسات ونماذج الاتصالات المدمجة في الكثير من الأشياء التي نستخدمها يوميًّا ــ تتعلق بشكل جزئي بالشبكات ولكنها تهتم بالقدر ذاته بتحويل كل ما يحيط بنا إلى بيانات.

بمجرد أن يتم تحويل العالم من حولنا إلى بيانات، ستتقيد الاستخدامات المحتملة لها فقط بالعبقرية البشرية. قام "موراي " بتحويل رحلات البحارة السابقين إلى بيانات من خلال جدولة يدوية طويلة ومؤلمة، وبالتالي تمكن من التوصل إلى أفكار وقيم رائعة. واليوم أصبحنا نمتلك الأدوات (الإحصاءات والمعادلات) والمعدات الضرورية (المعالجات والمساحات التخزينية الرقمية) لتنفيذ المهام المماثلة بصورة أسرع، وبحجم أكبر، وبصيغ مختلفة. في عصر البيانات الضخمة، حتى العثرات لها مميزاتها.

إننا بصدد إنشاء مشروع بنية تحتية عظيم ينافس في بعض أوجهه قنوات المياه الرومانية والتثقيف الذي أتاحته **الموسوعات**، ولكننا نفشل في تقدير هذه المشروعات لأن مشروعات هذا العصر جديدة للغاية، ولأننا في منتصف عملية إنشائها، ولأنها على العكس من المياه التي تنساب في القنوات المائية، فإن منتجات معاملنا غير ملموسة. المشروع هو التحويل البياني. مثله مثل غيره من تطويرات البنية التحتية، سيقوم هذا المشروع بإحداث تغييرات أساسية في المجتمع.

مكنت القنوات المائية من نمو المدن، وسهلت الطباعة من تثقيف الناس، ومكنت الصحف من نهوض الدولة القومية، ولكن كانت هذه البنى التحتية تهتم بالانسياب ــ الماء، المعرفة. وكذلك كان الهاتف والإنترنت. أما التحويل البياني فعلى النقيض؛

فهو يعرض إثراءً أساسيًّا للإدراك البشري. بمساعدة البيانات الضخمة، لن نواصل اعتبار عالمنا مجموعة من الأحداث التي نفسرها على أنها ظواهر طبيعية أو اجتماعية، ولكن على أنها كون كامل مكون من المعلومات.

لما يزيد على قرن من الزمان، افترض علماء الفيزياء أن هذه هي القضية ـــ ليست النواة هي أساس كل شيء بل المعلومات. ولكن عليّ أن أقر بأن الأمر لن يفهمه الكثير من الناس. بفضل التحويل البياني، أصبحنا الآن، في الكثير من الحالات، قادرين على جمع وحساب المعلومات بشكل أكثر شمولاً عن المظاهر المادية وغير المادية عن الوجود والعمل عليها.

إن رؤية العالم على أنه مكون من معلومات، كمحيطات من المعلومات يمكننا استكشافها على اتساع وعمق أكبر، قدمت لنا منظورًا عن الواقع لم نكن نمتلكه من قبل. إنه عبارة عن وجهة نظر عقلية يمكنها أن تخترق جميع مجالات الحياة. لقد أصبحنا الآن مجتمعًا إحصائيًّا لأننا أصبحنا نفترض أن العالم يمكن فهمه بواسطة الأعداد والرياضيات. وأصبحنا نأخذ إمكانية تحويل المعرفة عبر الزمان والمكان على أنه أمر مسلم به لأن فكرة الكلمة المكتوبة أصبحت راسخة في عقولنا. في المستقبل، قد تمتلك الأجيال التالية "وعي البيانات الضخمة" ـــ افتراض وجود مكونات كمية لكل ما نفعله، وأن هذه البيانات لا غنى عنها للمجتمع من أجل أن يتعلم منها. إن فكرة تحويل عشرات الآلاف من الأبعاد الواقعية إلى بيانات قد تبدو جديدة على سمع الكثيرين في الوقت الحاضر. ولكن في المستقبل، لا شك من أننا سنتعامل معها على أنها أمر مسلم به (الذي سيعود بنا إلى المصطلح الأصلي "البيانات").

في وقت ما، سيقلل تأثير التحويل إلى البيانات من تأثير كل من قنوات المياه والصحف، وربما قام بمنافسة الطباعة والإنترنت من خلال منحنا الوسائل لوضع رسم تفصيلي للعالم بطريقة قابلة للقياس والتحليل. في الوقت الحالي، يعتبر أكثر مستخدمي التحويل البياني تقدمًا هم رجال الأعمال؛ حيث يستخدمون البيانات الضخمة من أجل إنشاء أنواع جديدة من القيم المادية ـــ موضوع الفصل التالي.

6

القيمة

في أواخر تسعينيات القرن العشرين، تحولت شبكة الإنترنت بشكل سريع إلى مكان جامح وغير مرحب به وغير ودود، حيث كانت "برامج البريد الإلكتروني العشوائي" تغرق البريد الوارد للمستخدمين ومنتديات الإنترنت. في عام 2000، كان لدى "لويس فون أهن"، ذي الاثنين والعشرين ربيعًا, والذي كان قد تخرج للتو في الجامعة، حل لهذه المشكلة: إجبار المستخدمين على إثبات أنهم من البشر؛ لذا بحث عن أمر يسهل القيام به بواسطة البشر ولكنه يكون صعبًا على الآلات.

واتته فكرة عرض أحرف منحنية صعبة القراءة خلال عملية التسجيل في المواقع والخدمات، والتي يمكن للبشر أن يفكوا شفراتها ويكتبوا النص الصحيح في غضون ثوان، ولكن تعجز أجهزة الحاسب عن ذلك. نفذت شركة ياهو أسلوبه وتمكنت من التخلص من هجوم البريد العشوائي على الفور. أطلق "فون أهن" على ابتكاره اسم "كابتشاه: أمسكت بك" (اختصار للعبارة التالية في اللغة الإنجليزية: الاختبار الآلي العام للتفرقة بين البشر والحواسب). بعد مرور خمس سنوات، كان هذا الاختبار يطبق ملايين المرات يوميًّا.

منح هذا الابتكار "فون أهن" شهرة واسعة ووظيفة في تدريس علوم الحاسب في جامعة "كارنيجي ميلون" بعد أن حصل على درجة الدكتوراه، وكان أيضًا عاملاً مهمًّا في حصوله، في سن 27 عامًا، على واحدة من جوائز مؤسسة "ماك آرثر" للعباقرة التي تبلغ قيمتها نصف مليون دولار، ولكنه عندما أدرك أنه تسبب في إضاعة وقت ملايين الناس كل يوم من خلال كتابة الحروف المبهمة ــ كم كبير من المعلومات التي كانت تُمحى فيما بعد ــ لم يكن يشعر بأنه على هذا القدر من الذكاء.

عندما بحث عن طرق لاستخدام جميع هذه القوى الحاسوبية التي يمتلكها البشر في أمر مفيد، ابتكر البرنامج التالي، وأطلق عليه "ريكابتشاه". بدلاً من كتابة الناس حروفًا عشوائية، كان عليهم أن يكتبوا كلمتين من مشروعات نصوص ممسوحة ضوئيًّا لا يمكن لبرامج الحاسب البصرية التي تتعرف على الأحرف ــ لا يمكنها فهمها. إحدى الكلمتين تهدف إلى تأكيد ما كتبه المستخدمون وبالتالي تكون دليلاً على أن المستخدم من البشر، وأما الكلمة الثانية فهي كلمة جديدة بغرض إزالة الغموض عنها بالنسبة للآلات. للتأكيد على الدقة، يقدم البرنامج الكلمات غير الواضحة نفسها لخمسة مستخدمين مختلفين في المتوسط ليكتبوها بشكل صحيح قبل أن يؤكد صحتها. كان للبيانات في هذه الحالة استخدام رئيسي ــ إثبات أن المستخدم من البشر ــ ولكن كان هناك أيضًا غرض ثانوي: فك شفرات الكلمات غير الواضحة في النصوص المرقمنة.

كانت القيمة المادية التي كشف عنها هذا الابتكار هائلة، عندما يفكر المرء في تكلفة تعيين أشخاص ليقوموا بالمهمة ذاتها. عندما لا يتجاوز زمن الاستخدام في

المرة الواحدة 10 ثوان، ومع 200 مليون استخدام لهذا الابتكار كل يوم ــ المعدل الحالي ــ يصل المجموع إلى نصف مليون ساعة عمل يوميًّا. الحد الأدنى للأجور في الولايات المتحدة الأمريكية كان 7.25 دولار للساعة عام 2012. إذا ما كان المرء سيتحول إلى سوق رفع الغموض عن الكلمات التي لا يمكن للحواسب فهمها، فإن الأمر سيتكلف 4 ملايين دولار يوميًّا، أو أكثر من مليار دولار كل عام. ولكن بدلاً من هذا، قام "فون أهن" بتصميم نظام يقوم بالعمل ذاته مجانًا. كان هذا الأمر قيمًا للغاية عندما اشترت شركة جوجل هذه التقنية من "فون أهن" عام 2009. جعلت جوجل هذه التقنية متوافرة لأي موقع ليستخدمها، واليوم أصبح النظام مدمجًا في حوالي 200 ألف موقع إلكتروني من بينها فيس بوك وتويتر وكريجسليست.

تقلل قصة نظام "ريكابتشاه" أهمية إعادة استخدام البيانات. مع البيانات الضخمة، تتغير قيمة البيانات باستمرار. في العصر الرقمي، تخلت البيانات عن دورها الداعم للتعاملات التجارية وأصبحت هي نفسها السلعة التي يتم تبادلها. في عالم البيانات الضخمة، تتغير الأمور مرة بعد أخرى، حيث تتغير قيمة البيانات من استخدامها الأولي إلى استخداماتها المحتملة في المستقبل، وكانت لهذا الأمر نتائج مهمة، حيث أثر على أسلوب تقدير الشركات للبيانات التي تمتلكها وعلى اختيار من يمكنه الوصول إليها. وقد مكنت الشركات، وربما أجبرتها على أن تغير من أسلوب سير عملها. وغيرت من طريقة تفكير المؤسسات في البيانات وكيفية استخدامها.

كانت المعلومات على الدوام مهمة للتعاملات التجارية، حيث إنها، على سبيل المثال، تساعد على اكتشاف الأسعار المناسبة والتي تشير بدورها إلى كم الإنتاج. هذا البعد البياني مفهوم تمامًا. كانت هناك أنواع معينة من البيانات يتم تبادلها في الأسواق منذ زمن طويل، من بينها محتويات الكتب والمقالات والموسيقى والأفلام، والمعلومات المالية مثل أسعار الأسهم. في العقود القليلة الماضية، تم دمج هذه المعلومات مع المعلومات الشخصية. ملأ سماسرة البيانات في الولايات المتحدة الأمريكية مثل "أكسيوم" و"إكسبيريان"، و"إكويفاكس"، ملفات كاملة بالمعلومات الشخصية لمئات الملايين من المستخدمين. مع وجود مواقع التواصل الاجتماعي مثل فيس بوك وتويتر ولينكد إن، انضمت علاقاتنا الشخصية وآراؤنا ومفضلاتنا وأنماط حياتنا اليومية إلى الكم الكبير من المعلومات الشخصية المتوافرة عنا بالفعل.

باختصار، رغم أن البيانات كانت متوافرة طوال الوقت، فإنه كان يُنظر إليها إما على أنها وسيلة مساعدة للعمليات الجوهرية لإدارة الأعمال، أو أنها مقيدة لفئات محدودة نسبيًّا مثل حقوق الملكية الفكرية أو المعلومات الشخصية. على النقيض، في عصر البيانات الضخمة، ستكون جميع البيانات ذات قيمة بحد ذاتها.

عندما نقول "جميع البيانات"، فإننا نعني حتى الخام منها، التي يبدو معظمها معلومات تافهة. فكر في الحصول على معلومات من حساس حرارة مثبت على ماكينة في مصنع, أو الانسياب الفوري لإحداثيات نظام تحديد المواقع العالمي، أو قراءات مقياس السرعة، ومستويات استهلاك الوقود في عربات التوصيل ــ أو سرب مكون من 60 ألف عربة. أو فكر في المليارات من مصطلحات البحث

القديمة، أو أسعار جميع المقاعد على جميع رحلات الطيران التجارية في الولايات المتحدة الأمريكية لسنوات سابقة.

حتى وقت قريب، لم تكن هناك طرق سهلة لجمع تلك البيانات وتخزينها وتحليلها مما قيد بشكل كبير فرص استخراج قيمها المحتملة منها. في مثال "آدم سميث" عن صانع الدبابيس، والذي ناقش فيه قطاع العمال في القرن الثامن عشر، ستجد أن الأمر يتطلب مراقبين يتابعون جميع العمال ليس فقط من أجل دراسة واحدة، ولكن لجميع الأوقات كل يوم، مسجلين قياسات مفصلة ويدونون النتائج على أوراق سميكة بريشة حبر. عندما فكر علماء الاقتصاد القدامى في عوامل الإنتاج (الأرض والعمالة ورأس المال)، لم تكن فكرة ترويض البيانات موجودة. رغم أن تكلفة جمع البيانات واستخدامها قد تضاءلت كثيرًا في العقدين الماضيين، فإنها وحتى وقت قريب كانت تعتبر باهظة إلى حد ما.

إن ما يميز الحقبة التي نعيش فيها هو أن الكثير من العقبات الموروثة عن جمع البيانات لم تعد موجودة، حيث وصلت التكنولوجيا إلى حد جعل جمع كميات كبيرة من المعلومات وتسجيلها لا يحتاجان إلى كلفة كبيرة. يمكن أن يتم تجميع البيانات آليًّا، بدون حتى أي جهد أو إدراك من قبل من يتم تسجيل المعلومات عنهم. ولأن تكلفة تخزين البيانات قد انخفضت كثيرًا، فقد أصبح من السهل تقرير ما إذا كنا سنحتفظ بالبيانات بدلاً من التخلص منها. كل هذه الأمور وفرت كمًّا كبيرًا من البيانات لم يكن متوافرًا من قبل. على مدار نصف قرن مضى، كانت تكلفة التخزين الرقمي تقل إلى النصف كل عامين، في حين زادت السعة التخزينية إلى 50 مليون ضعف. طبقًا للشركات المعلوماتية مثل فايركاست وجوجل ــ حيث تدخل الحقائق الخام من أحد طرفي خط التجميع الرقمي وتخرج المعلومات المعالجة من الطرف الآخر ــ أصبحت البيانات تبدو كأنها المصدر أو العامل الجديد للإنتاج.

إن القيمة الفورية لأغلب البيانات تتضح لمن يجمعونها. في حقيقة الأمر، ربما كانوا يجمعونها لغرض معين، حيث تجمع المتاجر بيانات المبيعات من أجل القيام بحسابات مالية أكثر دقة. وتراقب المصانع منتجاتها النهائية من أجل التأكد من أنها تتوافق مع معايير الجودة. وتسجل المواقع الإلكترونية كل ضغطة يضغطها المستخدم ــ وأحيانًا حركة مؤشر الفأرة نفسها ــ من أجل تحليل المحتوى الذي تقدمه لزائريها وتعديله. تقرر هذه الاستخدامات الأولية للبيانات جمعها وتحليلها. عندما تسجل شركة أمازون ليس فقط الكتب التي يشتريها العملاء بل أيضًا الصفحات التي يزورونها، كانت تعلم أنها ستستخدم هذه البيانات من أجل عمل توصيات شخصية لهم. كذلك، تتبع فيس بوك "تحديثات الحالة" و"إعجاب" المستخدمين بموضوع ما من أجل تحديد ما تعرضه على صفحات موقعها من دعاية من أجل الحصول على المزيد من الأرباح.

على العكس من الأشياء المادية ــ الطعام الذي نتناوله والشمعة التي نضيئها ــ لا تقل قيمة البيانات عندما نستخدمها، حيث إنها يمكن معالجتها مرات عدة. المعلومات هي ما يطلق عليه علماء الاقتصاد، البضائع "خارج المنافسة": أي أن استخدام أحد الأشخاص لها لا يمنع مستخدمًا آخر من استخدامها. والمعلومات لا

تبلى بالاستخدام مثلما يحدث مع البضائع المادية. إذن، يمكن أن تستخدم شركة أمازون البيانات من تعاملات سابقة عندما تقوم بعمل توصيات جديدة لعملائها ــ وأن تستخدمها بشكل متكرر مع مستخدمين آخرين وليس فقط المستخدم الذي أنشئت من أجله.

الأمر الذي يعادل في أهميته استخدام البيانات مرات عدة للغرض ذاته، هو أنه يمكن توظيفها للعديد من الأغراض أيضًا. إن هذه النقطة مهمة للغاية عندما نحاول أن نفهم كم المعلومات التي ستكون قيمة بالنسبة لنا في عصر البيانات الضخمة. لقد رأينا بعضًا من هذه القدرات تتحقق بالفعل، عندما بحثت متاجر وولمارت في قواعد بيانات مبيعاتها القديمة واكتشفت العلاقة بين الأعاصير ومبيعات منتج بوب تارت.

كل هذه الأمور تؤدي بنا إلى أن القيمة الكاملة للبيانات تفوق بشكل كبير القيمة المستخرجة من استخدامها الأول. كما أنه يعني أن الشركات يمكنها أن تستخدم البيانات بشكل فعال حتى إن كان استخدامها الأول أو الاستخدامات التالية تعطي قيمًا ضئيلة، شريطة أن يواصلوا استخدام البيانات مرات عديدة.

"القيمة الاختيارية" للبيانات

لفهم ما تعنيه إعادة استخدام البيانات للحصول على قيمتها القصوى، فكر في السيارات الكهربائية، حيث يعتمد نجاحها كوسيلة من وسائل النقل على مجموعة كبيرة ومعقدة من النظم، التي تتعلق جميعها بمدى استمرار البطاريات الكهربية في الإمداد بالطاقة. يحتاج السائقون إلى إعادة شحن بطاريات سياراتهم بسرعة وبشكل كامل، وعلى شركات الطاقة أن تتأكد من أن الطاقة التي تستهلكها هذه السيارات لا تخل بأنظمتها. إننا نمتلك كمًّا كافيًا من محطات الوقود في الوقت الحالي، ولكننا لا نعلم حتى الآن احتياجات الشحن ولا أماكن وضع محطات الطاقة التي تحتاج إليها المركبات الكهربية.

الجدير بالذكر، أن هذه المشكلة لا تعتبر من مشكلات البنية التحتية بقدر كونها من المشكلات المعلوماتية. وتعتبر البيانات الضخمة من الأجزاء المهمة للتوصل إلى حل لها. في تجربة أجريت عام 2012، قامت شركة آي بي إم بالتعاون مع شركتي باسيفيك جاز وإليكتريك كومباني بولاية كاليفورنيا إلى جانب شركة هوندا للسيارات بجمع كم كبير من المعلومات للحصول على إجابة عن الأسئلة الأساسية المتعلقة بوقت ومكان احتياج السيارات الكهربية للحصول على الطاقة وما الذي سيعنيه هذا الأمر بالنسبة لمصادر الطاقة. قامت شركة آي بي إم بتطوير نموذج توقعات مدروس مبني على العديد من المدخلات: مستوى تخزين بطارية السيارة للطاقة، موقع السيارة، الوقت من اليوم، ومصادر الطاقة المتوافرة في محطة شحن الطاقة القريبة من موقع السيارة. نتج عن التجربة جمع لبيانات الاستهلاك الحالي للطاقة من الشبكة الكهربية وكذلك أنماط استهلاك الطاقة السابقة. عند تحليل هذا الكم الكبير من البيانات الحديثة والقديمة من مختلف المصادر، تمكنت آي بي إم من تحديد الأوقات والأماكن المحتملة التي سيحتاج السائقون إلى شحن بطاريات

سياراتهم فيها. كما أنها كشفت عن أفضل أماكن إنشاء محطات شحن السيارات. في نهاية الأمر، سيحتاج النظام إلى أن يضع في اعتباره اختلاف الأسعار في محطات الشحن القريبة من السيارة. هذا إلى جانب أن توقعات الطقس سيتم إدراجها في النظام هي الأخرى: على سبيل المثال، إذا كان الجو مشمسًا فستكون محطة الطاقة الشمسية القريبة ممتلئة بالطاقة، أما لو كانت توقعات الطقس تقول إن الأسبوع بأكمله سيكون مطيرًا فهذا يعني أن محطات الطاقة الشمسية ستكون متوقفة عن العمل.

إن النظام يستخدم المعلومات المجمعة من أجل غرض ما في غرض آخر ــ بمعنى آخر، تتحول البيانات من الاستخدام الأولي إلى الاستخدامات الثانوية، مما يزيد من قيمتها بمرور الوقت. إن مؤشر مستوى الطاقة بالبطاريات يخبر السائقين متى عليهم أن يعيدوا شحنها، ويتم جمع بيانات شبكة الطاقة بواسطة المؤسسة المسئولة ذاتها مما يمكنها من الحفاظ على استقرار الشبكة. هذه هي الاستخدامات الأولية. ويمكن إيجاد استخدامات ثانوية لهاتين المجموعتين من البيانات ــ قيم جديدة ــ عندما يتم تطبيقها على غرض مختلف تمامًا: تحديد متى وأين تحتاج السيارات لإعادة الشحن، وأين يجب إنشاء محطات إعادة الشحن. الأمر الأهم هو إدراج المعلومات الإضافية مثل مكان السيارة والاستهلاك السابق لشبكة الطاقة. تقوم شركة آي بي إم بمعالجة هذه البيانات ليس لمرة واحدة ولكن لمرات عدة في أثناء تحديثها المستمر لبرامج استهلاك السيارة الإلكترونية للطاقة وضغطها على شبكة الطاقة.

تشبه القيمة الحقيقية للبيانات جبال الجليد التي تطفو على سطح المحيطات، حيث لا يظهر إلا جزء ضئيل منها عند النظرة الأولى، في حين يختفي الجزء الأكبر منها تحت سطح الماء. يمكن للشركات المبتكرة التي تفهم هذا الأمر أن تستخرج القيم الخفية للبيانات وتحصد من ورائها أرباحًا ضخمة. باختصار، قيمة البيانات تحتاج إلى التفكير فيها من منطلق جميع الطرق المحتملة التي يمكن توظيفها فيها في المستقبل، وليس مجرد كيفية استخدمها في الوقت الحالي. لقد رأينا هذا الأمر في الكثير من الأمثلة التي عرضناها سابقًا. فقد وظفت شركة فايركاست استخدام البيانات من بيع تذاكر الطيران إلى توقع أسعارها، وأعادت شركة جوجل استخدام مصطلحات البحث للكشف عن تفشي وباء الأنفلونزا، وأعاد "موراي " استخدام سجلات قباطنة السفن القدامى للكشف عن تيارات المحيطات الصالحة للإبحار فيها.

ولكن لم يتم تقدير إعادة استخدام البيانات على الوجه الأكمل في عالم الأعمال وفي المجتمع حتى الآن. تمكن عدد محدود من المديرين التنفيذيين بشركة "كون إديسون" بمدينة نيويورك من أن يتخيلوا أن المعلومات عن الكابلات وسجلات الصيانة التي يبلغ عمرها قرنًا كاملاً من الزمان يمكن أن يتم استخدمها من أجل منع وقوع الحوادث في المستقبل. لقد تطلب الأمر جيلاً جديدًا من الإحصائيين وموجة جديدة من الأساليب والأدوات من أجل اكتشاف قيمة البيانات. حتى إن الكثير من شركات الإنترنت والتكنولوجيا لم تكن تدرك حتى وقت قريب قيمة إعادة استخدام

البيانات.

قد يكون من المفيد أن ننظر إلى البيانات بالطريقة نفسها التي ينظر بها علماء الفيزياء للطاقة، حيث يشيرون بمصطلحات "مخزنة" أو "محتملة" للطاقة المتواجدة في شيء ما ولكنها ساكنة. فكر في زنبرك مضغوط أو في كرة تقبع على قمة التل، حيث تكون الطاقة في هذه الأشياء كامنة ـــ محتملة ـــ حتى تنطلق، أو عندما يتم إطلاق العنان للزنبرك أو دحرجة الكرة من أعلى التل. في هذه الحالة تكون الطاقة في هذه الأشياء "نشطة" حيث إنها تتحرك وتؤثر على الأشياء الأخرى الموجودة في العالم من حولها. بعد استخداماتها الأولية، تظل قيمة البيانات متواجدة، ولكنها ساكنة، مخزنة قدرتها مثل الزنبرك أو الكرة، حتى يتم استخدام البيانات في استخدام ثانوي فتنطلق طاقتها الكامنة من جديد. وأخيرًا، في عصر البيانات الضخمة، أصبحنا نمتلك العقلية والأدوات اللازمة للحصول على القيمة الخفية داخل البيانات.

في النهاية، قيمة البيانات هي ما يمكن للمرء الحصول عليه من جميع الطرق المحتملة التي يمكن توظيف البيانات فيها. تبدو هذه الاستخدامات المحتملة اللانهائية مثل الخيارات ـــ ليس بمفهوم الأدوات المالية، ولكن بالمفهوم العملي للاختيار. إن قيمة البيانات هي مجموع هذه الخيارات: "القيمة الاختيارية" للبيانات. في الماضي، بمجرد أن يتم تحقيق الاستخدام الأساسي للبيانات، كنا نعتقد أنها أوفت بالغرض المطلوب منها، وكنا نقوم بمسحها أو تركها تذهب أدراج الرياح. ولكن في النهاية، بدا أن القيمة الأساسية فقط هي ما تم استخراجه من البيانات. في عصر البيانات الضخمة، تبدو البيانات مثل منجم الألماس السحري الذي يظل ينتج لفترة طويلة بعد انتهاء مخزونه الرئيسي. هناك ثلاث طرق لإطلاق القيمة الاختيارية للبيانات: إعادة الاستخدام الأساسية، ودمج المجموعات البيانية، وإيجاد "الحزم البيانية".

إعادة استخدام البيانات

من بين الأمثلة القديمة عن إعادة الاستخدام الابتكاري للبيانات هناك مصطلحات البحث. للوهلة الأولى، يبدو أن المعلومات غير ذات قيمة بعد انتهاء الغرض الرئيسي منها. والتفاعل اللحظي القائم يكون بين المستخدم ومحرك البحث الذي نتج عنه قائمة بالمواقع الإلكترونية والإعلانات التي قامت بمهمة معينة خاصة بهذه اللحظة فقط، ولكن يمكن أن تكون تلك المصطلحات القديمة قيمة للغاية. شركات مثل "هيتوايس"، وهي شركة لقياس الازدحام على شبكة الإنترنت والمملوكة لشركة سمسرة البيانات "إكسبيريان"، تساعد عملاءها على البحث المتعمق في ازدحام الإنترنت من أجل معرفة مفضلات المستهلكين. يمكن للمسوقين استخدام "هيتوايس" من أجل معرفة ما إذا كان اللون الوردي هو السائد في خطوط الموضة لهذا الربيع أم سيعود اللون الأسود مرة أخرى. أعدت شركة جوجل نسخة من برنامج تحليل مصطلحات البحث مفتوحًا للمستخدمين ليفحصوه. وقامت الشركة بإطلاق خدمة لتوقعات عالم الأعمال بالتعاون مع ثاني أكبر بنك في إسبانيا، "بي بي

في إيه"، لمتابعة قطاع السياحة وفي الوقت ذاته بيع مؤشرات اقتصادية فورية تعتمد على بيانات البحث. يستخدم بنك بريطانيا مصطلحات البحث المتعلقة بالعقارات من أجل معرفة ما إذا كانت أسعار المنازل ترتفع أم تنخفض.

تعلمت الشركات التي لم تقدر أهمية إعادة استخدام البيانات الدرس بالطريقة الصعبة، فعلى سبيل المثال، في بدايات شركة أمازون، أبرمت الشركة اتفاقًا مع شركة إيه أو إل لتدير التقنية الخاصة بموقع التجارة الإلكترونية الخاص بشركة إيه أو إل من الباطن. بدا هذا الأمر لأغلبية الناس أنه اتفاق عادي لإسناد العمل لشركة أخرى، ولكن ما كانت تهتم به شركة أمازون في الأساس، كما يشرح "أندرياس ويجيند"، كبير علماء شركة أمازون السابق، هو الحصول على بيانات عما يبحث عنه ويشتريه مستخدمو إيه أو إل، الأمر الذي كان سيحسن من أداء محرك التوصيات الخاص بها. ولم تدرك شركة إيه أو إل هذا الأمر قط، حيث إنها لم تر سوى قيمة البيانات في استخدامها الأولي فقط ـــ المبيعات. أما شركة أمازون فقد حصدت الكثير من الأرباح من جراء إعادة استخدام البيانات مرة أخرى.

أو كحالة برنامج شركة جوجل الخاص بالتعرف على الكلمات صوتيًّا تحت عنوان GOOG-411 لإعداد قوائم البحث العامة، والذي عمل منذ عام 2007 وحتى عام 2010. لم تكن عملاقة البحث على الإنترنت تمتلك تكنولوجيا التعرف على الكلمات صوتيًّا، لذا فقد احتاجت إلى إسناد إنتاجها لشركة أخرى. وقد توصلت إلى اتفاق مع شركة "نوانس" الرائدة في هذا المجال، والتي كانت متحمسة للعمل مع هذا العميل الكبير. ولكن شركة "نوانس" كانت جاهلة بماهية البيانات الضخمة: لم يذكر العقد من عليه أن يحصل على تسجيلات الترجمة الصوتية، وقد احتفظت بها شركة جوجل لنفسها. سمح تحليل هذه البيانات للمرء بأن يحدد احتمالية أن أحد الملفات الصوتية الرقمية قد يرتبط بكلمة بعينها. وكان هذا الأمر ضروريًّا لتطوير تكنولوجيا التعرف الصوتي على الكلمات أو ابتكار خدمة جديدة تمامًا. في هذا الوقت كانت شركة "نوانس" تعتبر نفسها رائدة مجال امتيازات البرمجيات، وليس معالجة البيانات. وبمجرد أن أدركت الخطأ الذي وقعت فيه، بدأت في إبرام اتفاقيات مع شركات شبكات المحمول ومنتجي سماعات الهاتف لتستخدم خدمتها للتعرف الصوتي على الكلمات ـــ وبهذا تمكنت من جمع البيانات.

إن قيمة إعادة استخدام البيانات أمر مفيد للمؤسسات التي تجمع أو تتحكم في كم كبير من البيانات ولكنها لا تستفيد بها، مثل الشركات التقليدية التي تعمل غالبًا خارج إطار شبكة الإنترنت، والتي قد تكون جالسة على منابع معلوماتية غير مطروقة. بعض الشركات ربما تكون قد جمعت البيانات واستخدمتها لمرة واحدة (إن استخدمتها)، وقامت بتخزينها بسبب قلة السعة التخزينية ـــ في "مقابر البيانات"، كما يطلق علماء البيانات على الأماكن التي تخزن فيها البيانات القديمة.

إن شركات الإنترنت والتكنولوجيا تقع في المقدمة فيما يتعلق باستئناس طوفان البيانات، حيث إنها تجمع كمًّا كبيرًا من البيانات فقط لأنها تعمل عبر الإنترنت وتسبق بقية المجالات في تحليلها. ولكن تركز جميع الشركات على الربح. أشار مستشارو شركة ماكينزي وشركاه إلى إحدى شركات النقل، لم يذكروا اسمها،

والتي لاحظت خلال عمليات توصيل البضائع وجود كميات كبيرة من المعلومات عن شحنات المنتجات في جميع أنحاء العالم. وانتهزت هذه الشركة الفرصة، وأنشأت قسمًا خاصًّا لبيع البيانات المتراكمة في شكل توقعات عملية واقتصادية. بمعنى آخر، أنها قامت بإنشاء نسخة خارج الإنترنت من عمل شركة جوجل في مجال تحليل مصطلحات البحث السابقة. أو يمكنك التفكير في "سويفت" نظام تحويل الأموال السلكي الدولي بين البنوك، والذي لاحظ أن التحويلات لها علاقة وثيقة بالنشاط الاقتصادي الدولي، لذا تمكن النظام من القيام بتوقعات للناتج المحلي لبعض الدول بناءً على الأموال التي يتم تحويلها من خلال شبكته.

بعض الشركات، بفضل موقعها في سلسلة قيمة المعلومات، قد تكون قادرة على جمع كم هائل من البيانات، رغم عدم الحاجة لها في الوقت الحالي أو لم تكن ماهرة في إعادة استخدامها. على سبيل المثال، تجمع شركات شبكات المحمول معلومات عن أماكن المشتركين فيها بحيث يمكنها أن تمرر المكالمات. بالنسبة لهذه الشركات، كان لهذه البيانات استخدامات تقنية محدودة، ولكنها أصبحت ذات قيمة أعلى عندما أعيد استخدامها بواسطة شركات تعمل على توزيع دعاية وعروض مخصصة معتمدة على مكان المستخدم. لا تنبع قيمة البيانات أحيانًا من النقاط البيانية الفردية بل مما ستكشفه عند تراكمها؛ لذا تتمكن شركات تحديد المواقع الجغرافية مثل إير سايج وسينس نتووركس التي ناقشناها في الفصل السابق من بيع المعلومات عن أماكن تجمع الناس في ليالي أيام الجمعة أو عن حالة الاختناقات المرورية. يمكن استخدام هذه المعلومات المكدسة من تحديد قيمة مادية حقيقية للعقارات أو أسعار لوحات الإعلانات.

حتى أتفه المعلومات قد تحتوي على قيمة خاصة إذا ما تم استخدامها بالصورة الصحيحة. فكر مرة أخرى في شركات شبكات المحمول, التي تمتلك معلومات عن أوقات وأماكن اتصال الهواتف بالشبكة الرئيسية وكذلك معلومات عن قوة الشبكة في هذه الأماكن. لطالما استخدمت شركات المحمول هذه المعلومات من أجل تحسين أداء شبكاتها وتقرير أين تجب إضافة أو تطوير البنى التحتية. ولكن لهذه البيانات الكثير من الاستخدامات المحتملة، حيث استخدمها مصنعو السماعات لمعرفة ما يؤثر على قوة الإشارة، على سبيل المثال، من أجل تحسين جودة الاستقبال لأجهزتها. لطالما عزفت شركات المحمول عن تحويل استخدام هذه البيانات خشية وقوعها في مشكلات تتعلق بالخصوصية. ولكنها أصبحت أكثر استعدادًا لتغيير موقفها وأصبحت تعتبر بياناتها مصادر محتملة للدخل. في عام ، قامت شركة المحمول الإسبانية العالمية العملاقة "تليفونيكا" بإنشاء شركة جديدة أطلقت عليها تليفونيكا للأفكار الرقمية لبيع بيانات عن أماكن المشتركين المجهولة والمتراكمة إلى أصحاب المتاجر وغيرهم.

البيانات المعاد دمجها

أحيانًا تنطلق القيمة الكامنة في البيانات من خلال دمج إحدى المجموعات البيانية مع مجموعة أخرى، وربما كانت مجموعة بيانية مختلفة تمامًا عن الأولى. يمكننا أن

نقوم بأمور مبتكرة من خلال المزج بين البيانات بطرق جديدة. من بين الأمثلة عن كيفية القيام بهذا الأمر دراسة ماهرة نُشرت عام 2011 عن احتمالية مساهمة الهواتف المحمولة في الإصابة بالسرطان. مع وجود حوالي ستة مليارات هاتف محمول في جميع أنحاء العالم، تقريبًا هاتف واحد لكل فرد من سكان الكرة الأرضية، كانت المسألة التي تتناولها تلك الدراسة مهمة للغاية. حاولت الكثير من الدراسات البحث عن صلة بين الأمرين، ولكنها أعيقت بسبب عيوب الأساليب البحثية، فقد كانت عينات البحث صغيرة للغاية أو أن المدى الزمني الذي كانوا يقومون بتغطيته قصير للغاية، أو كانوا يعتمدون على بيانات يجمعونها بأنفسهم وتمتلئ بالأخطاء. ولكن تمكن فريق من الباحثين من جمعية السرطان الدانمركية من ابتكار طريقة مشوقة تعتمد على البيانات التي تم جمعها من قبل.

إن البيانات عن جميع الهواتف المحمولة منذ إطلاقها في الأسواق في الدانمرك تم جمعها من شبكات المحمول. أجريت الدراسة على الأشخاص الذين امتلكوا الهواتف المحمولة منذ عام 1987 وحتى عام 1995، باستثناء المشتركين عن طريق المؤسسات أو الذين لم تتوافر بياناتهم الاقتصادية والاجتماعية. وقد توصل الفريق إلى عدد 358403 أشخاص. كانت الدولة بدورها تحتفظ بسجلات على مستوى الدولة بأكملها عن مرضى السرطان، والتي احتوت على عدد 10729 شخصًا أصيبوا بأورام في الجهاز العصبي المركزي في الفترة من 1990 وحتى عام 2007، فترة المتابعة. وأخيرًا استخدمت الدراسة السجل القومي الذي يحتوي على معلومات عن أعلى مستويات الدراسة والدخل المتاح لكل من المواطنين الدانمركيين. بعد مزج المجموعات البيانية الثلاث، ألقى الباحثون نظرة على ما إذا كان مستخدمو الهواتف المحمولة تظهر بينهم معدلات إصابة بالسرطان أعلى من غير المشتركين بالخدمة، وما إذا كان الأشخاص الذين يمتلكون الهواتف المحمولة لفترات طويلة أكثر عرضة للإصابة بالسرطان؟

رغم حجم الدراسة، لم تكن البيانات فوضوية أو غير دقيقة على الإطلاق: كانت المجموعات البيانية المطلوبة ذات معايير متناهية الجودة من أجل الأغراض الطبية أو التجارية أو الديموغرافية. لم تجمع المعلومات بطرق من شأنها تقديم اتجاهات تتعلق بسياق الدراسة. في الحقيقة، جُمعت هذه المعلومات قبل الدراسة بسنوات طويلة، ولأسباب لا تتعلق بهذه الدراسة. الأمر الأهم، أن الدراسة لم تكن تعتمد على عينة بحثية بل كانت تعتمد على معادلة ن = الجميع: جميع حالات الإصابة بالسرطان تقريبًا، وجميع مستخدمي الهواتف المحمولة تقريبًا، والذين وصلوا إلى حوالي 3.5 مليون شخص- بعد أعوام من بدء انتشار الهواتف المحمولة. كانت حقيقة احتواء الدراسة على جميع الحالات تقريبًا تعني أن الباحثين كانوا قادرين على التحكم في المجموعات الفرعية من المواطنين، مثل الأشخاص ذوي مستويات الدخل العالية.

في النهاية، لم تجد الدراسة أية زيادة في معدلات الإصابة بالسرطان تتعلق باستخدام الهواتف المحمولة. لهذا السبب لم تحدث هذه الدراسة أية ضجة إعلامية عندما نشرت في المجلة الطبية البريطانية *BMJ* عام 2011. ولكن إن وجدت أية

صلة بين الهواتف المحمولة والسرطان كانت الدراسة ستتصدر عناوين الأخبار في جميع أنحاء العالم، وكان سيتم تقدير أسلوب "البيانات المعاد دمجها" بدرجة كبيرة.

في البيانات الضخمة، يمتلك المجموع قيمة أكبر من أجزائه، وعندما نعيد دمج مجموعات من البيانات معًا، فإن قيمة المجموع الناتج بدورها ستكون أعلى من قيمة كل مجموعة على حدة. في هذه الأيام يعتبر مصطلح "تطبيقات الإنترنت المهجنة"، والذي يدمج مصدرين أو أكثر من مصادر البيانات لينتج شيئًا جديدًا مألوفًا على أسماع الكثير من المستخدمين. على سبيل المثال، يعمل موقع الممتلكات العقارية تحت عنوان "زيلو" على تركيب معلومات العقارات وأسعارها على خريطة لأحياء الولايات المتحدة الأمريكية. يقوم أيضًا بمعالجة الكثير من البيانات، مثل التعاملات الجديدة في كل حي ومواصفات كل عقار يتم تداوله، من أجل توقع قيمة بعض المنازل في منطقة بعينها. وقد سهل العرض الرسومي من إمكانية الحصول على البيانات، ولكن مع البيانات الضخمة يمكننا أن نفعل أكثر من هذا، فقد أمدتنا الدراسة الدانمركية بفكرة عن الاستخدامات المحتملة.

البيانات المرنة

من بين طرق إعادة استخدام البيانات هو تصميم مرونتها منذ البداية بحيث تكون ملائمة لاستخدامات متعددة. رغم أنه من غير الممكن تحقيق هذا الأمر طوال الوقت ــ حيث إن المرء قد يدرك الاستخدامات المحتملة للبيانات بعد مرور وقت طويل على جمعها ــ إلا أن هناك طرقًا متعددة لتشجيع الاستخدامات المتعددة للمجموعة البيانية نفسها. على سبيل المثال، بعض التجار يضعون كاميرات مراقبة في متاجرهم ليس فقط من أجل مراقبة اللصوص، بل أيضًا من أجل متابعة حركة البيع والشراء في المتجر وأين يقف المشترون ليشاهدوا المعروضات. يمكن للتجار أن يستخدموا هذه المعلومات من أجل وضع أفضل تصميم للمتاجر وللحكم على فاعلية حملات التسويق التي يقومون بها. كانت تستخدم هذه الكاميرات في السابق من أجل أغراض الأمن فقط، أما الآن فقد أصبح ينظر إليها على أنها استثمار من شأنه زيادة الدخل.

من بين أفضل حالات جمع البيانات مع وضع مرونتها في الاعتبار هي، كما هو متوقع، شركة جوجل، حيث تطوف سيارات الاستعراض الخاصة بها ــ التي أثارت الكثير من الجدل ــ جميع أنحاء العالم لتلتقط صورًا للمنازل والطرق، ولكنها تعمل أيضًا على محاكاة بيانات نظام تحديد المواقع العالمي، كما تفحص معلومات وضع الخرائط، حتى إنها تحتفظ بأسماء شبكات الإنترنت اللاسلكية (وربما المحتوى الموجود على الشبكات اللاسلكية مفتوحة المصدر، بشكل غير قانوني). إن جولة واحدة لإحدى هذه السيارات من شأنها أن تجمع عشرات الآلاف من التدفقات البيانية في كل لحظة. وجاءت المرونة في هذه الحالة من أن جوجل لم تستخدم البيانات بغرض استخدامها بطريقة أولية واحدة فقط، بل ستستخدمها في الكثير من الاستخدامات الثانوية أيضًا. على سبيل المثال، حسنت بيانات نظام تحديد المواقع العالمي التي جمعتها الشركة من خدمة الخرائط التي تقدمها الشركة

وكانت أساسية لتشغيل سيارتها ذاتية القيادة.

إن التكاليف الإضافية المتعلقة بجمع تدفقات متعددة من البيانات أو المزيد من النقاط البيانية في كل من هذه التدفقات البيانية عادة ما تكون منخفضة؛ لذا فمن المنطقي القيام بجمع أكبر كم ممكن من البيانات، وكذلك جعلها بيانات مرنة من خلال التفكير في استخدامات ثانوية محتملة لها منذ البداية، مما يزيد من القيمة الاختيارية للبيانات. الفكرة هي البحث عن "الحزم البيانية" ــ حيث تستخدم المجموعة البيانية الواحدة في حالات متعددة إذا كان من الممكن جمعها بطريقة معينة. أي أن البيانات قد تقوم بعمل مضاعف.

انخفاض قيمة البيانات

مع هبوط تكلفة تخزين البيانات الرقمية بشكل كبير، زادت محفزات الشركات الاقتصادية للاحتفاظ بالبيانات لإعادة استخدامها للأغراض نفسها أو لأغراض شبيهة، ولكن كانت هناك حدود لفائدتها.

على سبيل المثال، عندما استثمرت شركات مثل نتفليكس وأمازون أموالها في شراء المستخدمين لمنتجاتها والبحث على شبكة الإنترنت واستعراض توصيات المنتجات الجديدة، أغراها هذا الأمر باستخدام هذه السجلات مرات عدة طوال سنوات. مع وضع هذا الأمر في الاعتبار، قد يجادل المرء بأنه طالما لم تتقيد الشركة بالقيود القانونية والتشريعية مثل قوانين الملكية الفكرية، فإن عليها أن تستخدم السجلات الرقمية للأبد، أو على الأقل طالما كان هذا ممكنًا من الناحية الاقتصادية. ولكن لم يكن الواقع بهذه السهولة.

إن أغلب البيانات تفقد فائدتها بمرور الوقت. في مثل هذه الحالات، فإن استمرار الاعتماد على البيانات القديمة لا يعمل فقط على الفشل في إضافة قيمة جديدة لها، بل يعمل في واقع الأمر على تدمير قيمة البيانات الأحدث. على سبيل المثال، كتاب اشتريته من موقع أمازون منذ عشر سنوات، والذي أصبح لا يعكس اهتماماتك. إذا ما قامت أمازون باستخدام سجلات تعود أعمارها إلى عقد كامل مضى للقيام بتوصيات لكتب أخرى، فمن المحتمل ألا تشتري هذه الكتب ــ أو أن تهتم حتى بالتوصيات التالية التي قد يقدمها الموقع. ولأن توصيات موقع أمازون تعتمد على كل من البيانات القديمة والبيانات الحديثة، التي لا تزال تحتفظ بقيمتها، فإن وجود البيانات القديمة يقلل من قيمة البيانات الأحدث.

لذا امتلكت الشركة الدافع لاستخدام البيانات طالما أنها تحصل منها على فائدة. وبهذا كانت بحاجة دائمة إلى تمشيط بياناتها والتخلص من المعلومات التي فقدت قيمتها. يكمن التحدي في إدراك أن البيانات لم تعد مفيدة، ونادرًا ما يكون ترك اتخاذ هذا القرار معتمدًا على الوقت أمرًا صائبًا؛ لذا، قامت شركة أمازون وغيرها من الشركات بإنشاء نماذج دقيقة لمساعدتها على الفصل بين البيانات المفيدة والبيانات منتهية الصلاحية. على سبيل المثال، إذا ما كان أحد المستخدمين يستطلع كتابًا أوصى به الموقع بناءً على كتاب سابق قام بشرائه، فإن شركات التجارة الإلكترونية بإمكانها أن تستنتج أن الكتاب السابق الذي اشتراه المستخدم قد يظل

معبرًا عن مفضلاته الحالية. بهذه الطريقة يمكنهم أن يحددوا فائدة البيانات القديمة، وبالتالي يكون هذا النموذج دقيقًا في قياس "معدل انخفاض قيمة" المعلومات.

لا تنخفض قيمة جميع البيانات بالسرعة أو بالطريقة ذاتها، مما يوضح سبب اعتقاد بعض الشركات أنه يجب عليها أن تحتفظ بالبيانات لأطول وقت ممكن، حتى إن كانت السلطات التشريعية أو العامة يطالبون بمحو هذه البيانات أو جعلها مجهولة بعد فترة معينة من الوقت. على سبيل المثال، قاومت شركة جوجل كثيرًا المطالب بمحو عناوين بروتوكول الإنترنت الكاملة للمستخدمين من استفسارات البحث السابقة على محرك البحث الخاص بها. (إلا أنها قامت بدلاً من ذلك بمحو الأرقام الأخيرة فقط بعد تسعة أشهر من التجهيل التقريبي لاستفسارات البحث، حتى تتمكن الشركة من مقارنة البيانات من عام لآخر، مثل استفسارات أبحاث التسوق في أيام العطلات ـــ ولكن على أساس المناطق فقط وليس على مستوى الأفراد). هذا بالإضافة إلى أن معرفة أماكن الباحثين تساعد على تحسين النتائج المتعلقة بهم. على سبيل المثال، إذا ما كان هناك الكثير من الأشخاص يبحثون عن دولة تركيا ـــ ويدخلون إلى المواقع المتعلقة بهذه الدولة ـــ ستقوم المعادلات بوضع هذه الصفحات في ترتيب أعلى من الصفحات المتعلقة بمدينة نيويورك. حتى إن قلت قيمة البيانات في بعض أغراضها، قد تظل قيمتها الاختيارية عالية.

قيمة بواقي البيانات

قد يأخذ استخدام البيانات في بعض الأحيان شكلاً ماهرًا ومستترًا. يمكن لشركات الإنترنت أن تجمع البيانات عن جميع الأمور التي يقوم بها المستخدمون، وتعامل كل من هذه التعاملات على حدة كإشارة لاستخدامها مرجعية لتجعل الصفحات ملائمة لمفضلات المستخدمين، أو تحسين إحدى الخدمات أو إنشاء منتج رقمي جديد تمامًا. يمكننا أن نستعرض مثالاً قويًّا على هذا الأمر في قصة عن برنامجين للتدقيق الإملائي.

على مدار عشرين عامًا، طورت شركة مايكروسوفت برنامج تدقيق إملائي قويًّا لدمجه في برنامج وورد الخاص بها، حيث أنشأته من خلال مقارنة بين قاموس محدث باستمرار عن المصطلحات المكتوبة بشكل صحيح والأحرف التي يكتبها المستخدمون. أكد القاموس على الكلمات التي يدركها فقط، ويقوم النظام بمعاملة الاختلافات القريبة وغير الموجودة بالقاموس على أنها أخطاء إملائية يجب تصحيحها. بفضل الجهد المبذول من أجل تنظيم وتحديث القاموس، كان برنامج التدقيق الإملائي الخاص ببرنامج وورد من شركة مايكروسوفت متوافرًا فقط لأكثر اللغات انتشارًا، حيث إنه كلف الشركة ملايين الدولارات من أجل إنشائه والمحافظة عليه محدثًا.

فكر الآن في شركة جوجل والتي تمتلك أكثر برامج التدقيق الإملائي كمالاً في العالم، والذي يحتوي بشكل أساسي على جميع اللغات الحية في العالم. دائمًا ما يتم تحسين هذا النظام وإضافة كلمات جديدة له ـــ نتيجة عارضة لاستخدام الناس

لمحرك البحث يوميًّا. هل تخطئ في كتابة كلمة "آي باد"؟ ستجد تصحيحها. وماذا عن المصطلح الحديث الذي يشير إلى"قانون الصحة الذي وضعه الرئيس أوباما" Obamacare؟ ستجده أيضًا. هل فهمت؟

علاوة على ذلك، يبدو أن شركة جوجل قد حصلت على مدققها الإملائي بالمجان من خلال إعادة استخدام الأخطاء الإملائية التي يكتبها المستخدمون في محرك البحث المملوك للشركة في حوالي ثلاثة مليارات من الاستفسارات البحثية التي يجريها محرك البحث يوميًّا. تقوم حلقة المرجعيات الماهرة بإخبار النظام بالكلمات الصحيحة التي يقصد الناس أن يكتبوها. وأحيانًا ما "يعطي" المستخدمون الإجابة الصحيحة لمحرك البحث عندما يظهر السؤال الاستفساري في أعلى صفحة النتائج ـــ "هل تعني علم الأوبئة؟" ـــ من خلال الضغط على السؤال لبدء بحث آخر عن المصطلح الجديد. أو عندما يفتح المستخدمون صفحات تحتوي على الكلمات ذات التهجئة الصحيحة حيث إنها تتعلق بالكلمات المكتوبة بالشكل الصحيح أكثر من الكلمات المكتوبة بطريقة خاطئة. (إن هذا الأمر على قدر من الأهمية أكبر مما يبدو عليه: مع تحسن المدقق الإملائي من جوجل بشكل مستمر، توقف الناس عن مضايقة أنفسهم بالعناية بكتابة مصطلحاتهم البحثية بشكل صحيح حيث إن محرك البحث من جوجل سيتمكن من إعطائهم النتائج الصحيحة بغض النظر عن كتابتها بشكل صحيح أو خاطئ).

أظهر نظام التدقيق الإملائي من جوجل أن البيانات "السيئة" و"الخاطئة" و"المعيبة" قد تظل مفيدة للغاية. الجدير بالذكر أن شركة جوجل لم تكن الشركة الأولى التي واتتها تلك الفكرة. في حوالي عام 2000، رأت شركة ياهو إمكانية إنشاء مدقق إملائي من مصطلحات البحث المكتوبة بشكل خاطئ التي يكتبها المستخدمون في محرك البحث، ولكنهم لم يستفيدوا بهذه الفكرة قط. كانت بيانات مصطلحات البحث القديمة تعامل على أنها قمامة. وبالمثل، شركتا إنفوسيك وألتافيستا اللتان كانتا تعتبران محركات البحث الأكثر شهرة فيما سبق، كانت كل منهما تمتلك أكثر المجموعات البيانية عن الكلمات المكتوبة بشكل خاطئ في أيامهما، ولكنهما لم تتمكنا من تقدير قيمتها، حيث كانت أنظمتهما التي لم تكن ظاهرة للمستخدمين في أثناء عملها تعامل الكلمات المكتوبة على أنها "مصطلحات ذات صلة" وتكمل البحث. ولكنها كانت تعتمد على القواميس التي تخبر النظام بوضوح بأن هذه الكلمات صحيحة، وليس على تعاملات المستخدمين الحية المباشرة.

وحدها شركة جوجل هي من أدرك أن تفتيت تفاعلات المستخدمين يعتبر غبارًا من الذهب يمكن جمعه وصياغته ليصبح سبيكة ذهبية براقة. أحد أفضل مهندسي جوجل قدر أن برامج التدقيق الإملائي المقدمة من الشركة تعمل بشكل أفضل من برامج مايكروسوفت بقدر كبير (رغم أنه بعد أن تم الضغط عليه اعترف بأنه لم يقم بدراسة الأمر بطريقة يعتمد عليها). وسخر من فكرة أن هذه البرامج قد تم تطويرها "بالمجان"، حيث إن المادة الخام ـــ الكلمات المكتوبة بتهجئة خاطئة ـــ ربما تكون قد حصلت عليها الشركة دون تكلفة مباشرة، ولكن ربما أنفقت جوجل أموالاً أكثر

من التي أنفقتها مايكروسوفت من أجل تطوير النظام، قال هذا الأمر وقد ارتسمت على شفتيه ابتسامة كبيرة.

إن الطرق المختلفة التي اتبعتها الشركتان تخبرنا بالكثير، حيث رأت مايكروسوفت قيمة برنامج التدقيق عن طريق استخدامه غرضًا واحدًا فقط، معالجة النصوص. وعلى النقيض أدركت جوجل فائدة أعمق لهذا النظام، حيث لم تستخدم الشركة مفردات البحث من أجل تطوير أفضل برنامج تدقيق إملائي في العالم والحفاظ عليه محدثًا فحسب، بل طبقت هذا النظام أيضًا على الكثير من الخدمات الأخرى مثل سمة "الاستكمال الآلي" في محرك البحث وبريد جي ميل الإلكتروني ونصوص جوجل وحتى نظام الترجمة الذي قدمته الشركة.

ظهر مصطلح متخصص ليصف الطريق الرقمي الذي يتركه الناس خلفهم على شبكة الإنترنت: "بواقي البيانات"، والذي يشير إلى البيانات المستخدمة كمنتج ثانوي لأفعال وتحركات الناس في العالم. وفي عالم الإنترنت، يصف المصطلح تفاعلات المستخدمين عبرها: الأزرار التي يضغطونها وكم من الوقت يتصفحون الصفحات وأين يجول مؤشر الفأرة وماذا يكتبون وغيرها الكثير. صممت الكثير من الشركات أنظمتها بحيث تتمكن من جمع بواقي البيانات وإعادة تدويرها لتحسين خدمة حالية أو ابتكار خدمات جديدة. إن جوجل هي الملكة المتوجة في هذا المجال، حيث إنها تطبق مبدأ تكرار "التعلم من البيانات" على الكثير من الخدمات التي تقدمها. كل فعل يقوم به المستخدمون يعتبر إشارة ليتم تحليلها ووضعها كمرجعية في النظام.

على سبيل المثال، تدرك جوجل تمامًا عدد المرات التي يبحث فيها الناس عن مصطلح ما وكذلك المصطلحات المتعلقة به، وكم من المرات قاموا بالضغط على أحد الروابط ومن ثم عادوا مرة أخرى إلى صفحة البحث وهم يشعرون بالخيبة مما وجدوه ويبدءون في البحث من جديد، فهي تدرك ما إذا كان المستخدمون قد ضغطوا على الرابط الثامن من الصفحة الأولى أو الرابط الأول في الصفحة الثامنة ـــ أو ما إذا كانوا قد تركوا البحث بأكمله. ربما لم تكن الشركة هي الأولى التي راودتها هذه الفكرة، ولكن هي الوحيدة التي طبقتها بفاعلية لا مثيل لها.

قد تكون هذه المعلومات ذات قيمة عالية. إذا ما قام المستخدمون بالضغط على نتائج البحث المتواجدة في أسفل صفحة النتائج، فهذا يعني أنها ذات صلة بموضوع البحث أكثر من تلك المتواجدة في أعلى الصفحة، وتقوم معادلات تصنيف جوجل تلقائيًّا بوضعها في تصنيف أعلى من الصفحة في مرات البحث التالية. (وهي تفعل ذلك مع الإعلانات أيضًا). يقول أحد العاملين في جوجل: "إننا نهوى التعلم من المجموعات البيانية الضخمة الصاخبة".

إن بواقي البيانات هي الآلية التي تعمل خلف الكثير من الخدمات مثل التعرف الصوتي ومرشحات البريد العشوائي وترجمة اللغات وغيرها الكثير. عندما يشير المستخدمون إلى أن أحد برامج التعرف الصوتي قد فهم ما يقولونه بشكل خاطئ، فإنهم في حقيقة الأمر "يدربون" النظام على التحسن.

بدأت الكثير من الشركات في تصميم أنظمتها لتقوم بجمع المعلومات

واستخدامها بهذه الطريقة. في الأيام الأولى لموقع فيس بوك، فحص "مهندس البيانات" الأول بالموقع "جيف هامر باشر" (من بين الأشخاص الذين أسهموا في صياغة المصطلح) الكميات الكبيرة من بواقي البيانات. ووجد مع فريقه أن التوقع الرئيسي بأن الأشخاص قد يقومون بفعل ما (نشر مشاركة أو الضغط على أيقونة أو غيرها) يكون من خلال ما إذا كانوا قد رأوا أصدقاءهم يقومون بالفعل ذاته. لذا قام موقع فيس بوك بإعادة تصميم نظامه من أجل جعل أنشطة الأصدقاء أكثر وضوحًا، الأمر الذي تسبب في إنشاء حلقات جديدة من المشاركة في الموقع.

تكمن الفكرة في التوسع إلى ما يتخطى قطاع الإنترنت لأية شركة تمتلك طريقة لجمع تغذيات راجعة من المستخدمين. برامج قراءة الكتب الإلكترونية على سبيل المثال، يمكنها جمع كميات هائلة من البيانات عن المفضلات الأدبية وعادات الأشخاص الذين يستخدمونها: كم من الوقت يستغرقون في قراءة صفحة أو قسم، أين يقرأون، يقلبون الصفحات بعد قراءتها قراءة سطحية سريعة أم يغلقون الكتاب ولا يعودون لقراءته. تسجل الأجهزة أيضًا كل مرة يضع فيها القراء خطًّا تحت قطعة أدبية أو يدونون أمرًا ما في الهوامش. إن القدرة على جمع هذه النوعية من المعلومات تحوِّل القراءة، الفعل الذي كان فيما مضى فرديًّا، إلى نوع من الخبرة العامة.

بمجرد أن تتكدس البيانات، يمكن لبقايا البيانات أن تخبر الناشرين والمؤلفين بأمور لم يكونوا يعرفونها من قبل بالطرق القابلة للقياس: الكتب التي أعجبت القراء والكتب التي لم تعجبهم، وأنماط قراءة الناس.

إن هذه المعلومات قيمة من الناحية التجارية، حيث يمكن أن يتخيل المرء أن تبيع شركات الكتب الإلكترونية هذه المعلومات للناشرين من أجل تحسين محتوى الكتب وبنيتها. على سبيل المثال، كشف تحليل شركة بارنز ونوبل للبيانات التي جمعتها من قارئ الكتب الإلكترونية الخاص بها الذي يطلق عليه اسم نووك ــ كشف أن الناس يميلون إلى التوقف عن قراءة الكتب الطويلة غير الخيالية من منتصفها، حيث ألهم هذا الاكتشاف الشركة بأن تنتج مجموعة أطلقت عليها "قصاصات نووك": أعمال قصيرة عن موضوعات مثل الصحة والموضوعات الحالية.

أو فكر في برامج التعليم عبر الإنترنت مثل أودا سيتي وكور سيرا وإيدكس، حيث تتابع هذه البرامج التفاعلات التي تحدث بين الطلبة عبر الإنترنت من أجل التوصل إلى أفضل ما يصلح لهم من الناحية التربوية. وصلت الفصول الدراسية لهذه البرامج إلى عشرات الآلاف من الطلبة، الأمر الذي ينتج كمًّا هائلاً من البيانات. فيمكن للأساتذة في الوقت الحالي أن يعرفوا ما إذا قامت نسبة كبيرة من الطلبة بإعادة مشاهدة جزء معين من محاضراتهم، الأمر الذي يفترض أنهم لم يتمكنوا من إيضاح تلك النقطة جيدًا. عندما كان يدرس دورة تعليمية في برنامج كور سيرا للتعليم الإلكتروني، أشار الأستاذ بجامعة ستانفورد "أندرو إن جي" إلى أن حوالي 2(طالب فهموا سؤالاً معينًا من واجباتهم المدرسية بشكل خاطئ ــ وأجابوا بالإجابة الخاطئة ذاتها. وبدا أنهم قد وقعوا جميعهم في الخطأ ذاته. ولكن ما الذي حدث

بالفعل؟
مع القليل من البحث، وجد أنهم قاموا بعكس معادلتين جبريتين في معادلة لوغاريتمات؛ لذا في الوقت الحالي عندما يقع طلبة آخرون في الخطأ ذاته، لم يعد النظام يقول إنهم مخطئون وحسب، بل يعطيهم تلميحًا لمراجعة حساباتهم. يطبق النظام البيانات الضخمة أيضًا، حيث إنه يقوم بتحليل مشاركات المنتدى التي قرأها الطلبة وما إذا كانوا قد أكملوا واجباتهم المنزلية بالشكل الصحيح ليتوقع احتمالية ما إذا كان الطلبة الذين يقرءون مشاركة معينة سيتمكنون من الإجابة الصحيحة على الأسئلة أم لا، وهي طريقة لتحديد أي مشاركات المنتدى تكون الأكثر فائدة للطلبة. لم يكن من الممكن معرفة هذه الأشياء من قبل، والتي من شأنها تغيير أسلوب التعليم والتعلم إلى الأبد.

قد تكون بقايا البيانات من أكثر المميزات التنافسية التي تمتلكها الشركات. وقد تصبح أيضًا عائقًا ضخمًا في وجه المنافسين. فكر: إذا ما أطلقت إحدى الشركات حديثة العهد بالعمل موقعًا للتجارة الإلكترونية، أو موقعًا للتواصل الاجتماعي أو محرك بحث أفضل من محركات البحث الرائدة في الوقت الحالي مثل أمازون أو جوجل أو فيس بوك، فقد يكون من الصعب عليها المنافسة في السوق الحالية، ليس بسبب حجم استثماراتها وتأثير شبكة الإنترنت أو العلامة التجارية، ولكن بسبب أن أداء الكثير من هذه الشركات الرائدة يعتمد على بقايا البيانات التي تجمعها من تفاعلات المستخدمين على الإنترنت وأنها تعيد إدراجها مرة أخرى في خدماتها. هل يمكن لموقع تعليم إلكتروني جديد أن يدرك كيفية منافسة موقع آخر يمتلك بالفعل كمًّا هائلاً من البيانات والتي يمكنها أن تجعله يدرك الأمور التي تصلح للطلبة على الوجه الأمثل؟

قيمة البيانات مفتوحة المصدر

من المحتمل أن نفكر في الوقت الحالي في مواقع مثل جوجل وأمازون على أنها رائدة البيانات الضخمة، ولكن لا شك في أن الحكومات هي الجامع الأساسي للمعلومات على المقياس الكبير، وما زالت حتى وقتنا هذا تنافس أية مؤسسات خاصة في كم البيانات التي تتحكم فيها. هناك اختلاف واحد بين الحكومات ومالكي البيانات في القطاع الخاص هو أن الحكومات عادة ما تجبر الناس على إمدادها بالمعلومات، بدلاً من أن تقنعهم بفعل هذا الأمر أو أن تعرض عليهم شيئًا ما في المقابل. ونتيجة لهذا ستواصل الحكومات الحصول على كميات هائلة من البيانات.

إن الدروس المستفادة من البيانات الضخمة تنطبق على القطاع العام مثلما تنطبق على الكيانات التجارية: إن قيمة البيانات الحكومية تكون كامنة وتطلب طرقًا تحليلية مبتكرة لإطلاقها. ولكن رغم الوضع الخاص الذي تتمتع به الحكومات فيما يتعلق بجمع المعلومات، فإنها لا تتمتع بالفاعلية فيما يتعلق باستخدامها. ولكن برزت مؤخرًا فكرة تتعلق بأن أفضل طريقة لاستخراج القيمة من البيانات الحكومية هي إعطاؤها للقطاع الخاص أو المجتمع بوجه عام ليحاول ذلك. وكان هناك مبدأ يكمن خلف هذه الفكرة، فعندما تجمع الدولة البيانات، فإنها تفعل ذلك

بالنيابة عن مواطنيها، وبالتالي يجب أن يمتلك المجتمع قدرة الوصول إليها (باستثناء عدد قليل من الحالات، مثل الحالات التي تضر بالأمن القومي أو حقوق خصوصية الآخرين).

أدت هذه الفكرة إلى وجود عدد غير محدود من "البيانات الحكومية مفتوحة المصدر" في جميع أنحاء العالم. من منطلق أن الحكومات ما هي إلا أوصياء على المعلومات التي تجمعها، وأن القطاع الخاص والمجتمع قد يكونان أكثر ابتكارًا، طالب المناصرون للبيانات مفتوحة المصدر الكيانات الرسمية بأن تطلق البيانات بشكل علني من أجل استخدامها في الأغراض المدنية والتجارية. ولكي ينجح هذا الأمر، لا شك في أن البيانات يجب أن تكون في صورة معيارية يمكن لأجهزة الحاسب قراءتها لكي يمكن معالجتها بسهولة، وإلا فستكون البيانات عامة اسمًا فقط.

حظيت فكرة البيانات الحكومية مفتوحة المصدر بدفعة كبيرة عندما قام الرئيس "باراك أوباما"، في أول يوم عمل كامل له في 21 يناير 2009، بإصدار مذكرة رئاسية تأمر رؤساء الوكالات الفيدرالية بإطلاق أكبر كمّ ممكن من البيانات، حيث قال: "في وجه الشك ينتصر الانفتاح". وقد كان هذا الإعلان متميزًا، خاصة عند مقارنته بسلوك من سبقوه، الذين أمروا الوكالات الفيدرالية بالقيام بالعكس. أدى قرار "أوباما" إلى التعجيل بإنشاء موقع إلكتروني أطلقوا عليه data.gov، مخزن للمعلومات مفتوحة المصدر التي يمكن الحصول عليها من الحكومات الفيدرالية. تزايد كم البيانات في هذا الموقع من 47 مجموعة بيانية عام 2009 إلى حوالي 450 ألفًا من 172 وكالة فيدرالية بعد مرور ثلاث سنوات في عام 2012.

حتى في بريطانيا الحريصة على معلوماتها، والتي قد تم احتجاز الكثير من المعلومات الحكومية بواسطة حقوق النشر والملكية التي من الصعب والمكلف جدًّا الحصول على تصريح باستخدامها (مثل الرموز البريدية لشركات التجارة الإلكترونية)، لاح في الأفق بُعد من التقدم الجزئي، حيث أصدرت الحكومة في المملكة المتحدة قوانين تشجع على المعلومات مفتوحة المصدر وعززت هذا بإنشاء معهد البيانات مفتوحة المصدر الذي يشارك في إدارته "تيم بيرنرزلي"، مخترع الشبكة العنكبوتية العالمية، للتشجيع على ابتكار استخدامات مبتكرة للبيانات مفتوحة المصدر وطرق لتحريرها من قبضة الدولة.

أعلن الاتحاد الأوربي هو الآخر عن بعض المبادرات للبيانات مفتوحة المصدر التي قد تشمل كامل القارة الأوروبية في وقت قريب. قامت دول أخرى من أماكن أخرى في العالم، مثل أستراليا والبرازيل وشيلي وكينيا، بإصدار إستراتيجيات البيانات مفتوحة المصدر وتنفيذها. وعلى المستوى تحت القومي، يقوم عدد متزايد من المدن والبلديات من جميع أنحاء العالم هي الأخرى بتطبيق البيانات مفتوحة المصدر، وكذلك فعلت المؤسسات الدولية مثل البنك الدولي، الذي وفر المئات من المجموعات البيانية التي كانت سرية في الماضي عن الاقتصاد والمؤشرات الاجتماعية.

في الوقت ذاته، استكشف مطورو الإنترنت والمفكرون من أصحاب الرؤى

الثاقبة البيانات من أجل إيجاد طرق للاستفادة القصوى منها، مثل مؤسسات كود فور أمريكا ومؤسسة صن لايت في الولايات المتحدة الأمريكية ومؤسسة المعرفة المنفتحة في المملكة المتحدة. من بين الأمثلة الأولى للاحتمالات التي يمكن الحصول من خلالها على البيانات مفتوحة المصدر موقع يسمى FlyOnTime.us، حيث يمكن لزوار هذا الموقع أن يكتشفوا بفاعلية (من بين الكثير من الأمور الأخرى) ما إذا كان الطقس العاصف سيتسبب في تأخير الرحلات الجوية في مطار معين. يقوم الموقع بالمزج بين معلومات الرحلات الجوية والطقس التي يحصل عليها من مصادر البيانات الرسمية المتوافرة مجانًا ويمكن الحصول عليها من خلال شبكة الإنترنت. لقد تم إنشاء هذا الموقع بواسطة المؤيدين للبيانات مفتوحة المصدر ليظهروا فائدة المعلومات التي تجمعها الحكومات الفيدرالية. حتى إن البرنامج الإلكتروني المستخدم مفتوح المصدر يمكن للآخرين أن يتعلموا منه ويعيدوا استخدامه.

يترك موقع FlyOnTime.us البيانات تقوم بالعمل كله، وعادة ما تقوم البيانات بأمور مذهلة. قد يرى المرء فيما يتعلق بالرحلات من بوسطن إلى نيويورك أن المسافرين بحاجة إلى أن يكونوا مستعدين لتأخر الرحلات لفترتين بسبب الضباب أكثر من الثلج. لم يكن هذا ما توقعه المسافرون في أثناء وجودهم في صالة المغادرة، حيث كانوا يعتقدون أن السبب الرئيسي لتأخر الرحلة هو الثلج. ولكن هذا هو نوع المعرفة الذي تجعله البيانات الضخمة ممكنًا، فعندما يقوم شخص ما بمعالجة البيانات السابقة من مكتب النقل عن تأخر الرحلات إلى جانب معلومات المطار الحالية من إدارة الطيران الفيدرالية، مع تقارير الطقس السابقة من الإدارة الوطنية لدراسة المحيطات والغلاف الجوي ومعلومات عن الحالة الجوية الحالية من خدمة الطقس الوطنية. يوضح الموقع كيف يمكن لكيان لا يمكنه جمع التدفقات المعلوماتية أو التحكم بها، مثل محركات البحث أو كبار التجار، أن يظلوا قادرين على امتلاك البيانات واستخدامها لاستخراج قيمتها الكامنة.

تقييم ما لا يقدر بثمن

سواء كانت منشورة للعامة أو محفوظة في سراديب المؤسسات، من الصعب قياس قيمة البيانات. تذكر الأحداث التي وقعت في يوم الجمعة 18 مايو 2012. في هذا اليوم، قام مؤسس موقع فيس بوك "مارك زوكربيرج" الذي كان يبلغ من العمر 28 عامًا رمزيًّا بقرع ناقوس افتتاح مؤشر ناسداك من المقر الرئيسي للشركة في ميلانو بارك بكاليفورنيا، حيث بدأ موقع التواصل الاجتماعي الأكبر في العالم ـــ الذي يبلغ عدد مستخدميه شخصًا واحدًا من بين كل عشرة أشخاص في العالم ـــ حياته الجديدة كشركة عامة. زادت أسهم الشركة على الفور بنسبة 11 %، كما يحدث مع الكثير من الأسهم التكنولوجية في اليوم الأول للتعاملات بها. ولكن حدث أمر غريب فيما بعد، حيث بدأت أسهم شركة فيس بوك في الانخفاض. لم تساعد حقيقة وجود خلل في أجهزة حاسب ناسداك في توقف التعاملات بشكل مؤقت، فقد كانت المشكلة الحقيقية تحدث. عندما شعروا بالخطر، قام وكلاء الأسهم،

الذين يقودهم "مورجان ستانلي"، بالحفاظ على قائمة الأسهم بحيث تظل أعلى من السعر الأصلي لإصدارها.

في الليلة السابقة لما حدث، حددت البنوك التي تتعامل مع شركة فيس بوك سعر السهم للشركة بقيمة 38 دولارًا للسهم، أي حوالي 104 مليارات دولار. (الذي من باب المقارنة، يعادل رأس المال السوقي لشركة بوينج وجنرال موتورز وديل للحواسب مجتمعة). ما القيمة الفعلية لشركة فيس بوك؟ في مراجعة حسابات الشركة المالية لعام 2011، الذي من خلاله خمن المستثمرون قيمة الشركة، قدمت الشركة تقارير عن أصول ثابتة تبلغ قيمتها 6.3 مليار دولار، وهي قيمة أجهزة الحواسب ومعدات المكتب وغيرها من الأشياء المادية. كم بلغت القيمة الدفترية للكميات الهائلة من المعلومات المخزنة في سراديب مؤسسة فيس بوك؟ لا شيء. لم تكن مدرجة رغم أن الشركة لم تكن أي شيء سوى بيانات.

أصبح الموقف أكثر غرابة فيما بعد، حيث قام "دوج لاني"، نائب رئيس قسم الأبحاث بشركة "جارتنر"، إحدى شركات أبحاث السوق، بمعالجة الأرقام للفترة السابقة لطرح الشركة للاكتتاب العام وتوصل إلى أن شركة فيس بوك تمكنت من جمع 2.1 تريليون قصاصة من "المحتوى القابل إلى التحويل إلى أموال" في الفترة ما بين عامي 2009 و2011، مثل "الإعجاب"، والمواد المنشورة والتعليقات. بالمقارنة بتقييم الاكتتاب العام، يعني أن كل قصاصة من هذا المحتوى، عند اعتبارها نقطة بيانية مستقلة، تبلغ قيمتها 5 سنتات. بنظرة أخرى على الأمر، تبلغ قيمة كل مستخدم من مستخدمي فيس بوك حوالي 100 دولار، حيث إن المستخدمين هم مصدر المعلومات التي يجمعها موقع فيس بوك.

كيف يمكن تفسير الاختلاف بين قيمة فيس بوك بموجب المعايير المحاسبية (6.3 مليار دولار) والقيمة المبدئية في السوق (104 مليارات دولار)؟ لا توجد طريقة جيدة للقيام بهذا الأمر. ولكن هناك اتفاقًا عامًّا على أن الطريقة الحالية لتقدير قيمة المؤسسات، من خلال الاستعانة "بالقيمة الدفترية" للشركة (الذي يكون في الغالب المبالغ المالية المملوكة للشركة والأصول المادية)، لم تعد تعكس القيمة الحقيقية للمؤسسات. في حقيقة الأمر، كانت الفجوة بين القيمة الدفترية و"القيمة السوقية" للشركات ـــ التي يمكن أن تكتسبها الشركة من سوق الأوراق المالية أو من خلال شرائها من مالك جديد ـــ تتسع على مدى عقود طويلة. حتى إن مجلس الشيوخ الأمريكي قد عقد جلسات استماع عام 2000 بغرض تحديث قواعد التقارير المالية، التي ظهرت في ثلاثينيات القرن العشرين عندما كانت الشركات القائمة على المعلومات نادرة الوجود. لقد أثرت هذه المشكلة على أكثر من مجرد تقرير الموازنة للشركة: عدم القدرة على تقدير قيمة الشركات بالشكل الصحيح من شأنه أن يتسبب في المخاطرة بالشركات وعدم ثبات السوق.

إن الاختلاف بين القيمة الدفترية للشركة وقيمتها السوقية يحسب على أنه "الأصول غير الملموسة"، والتي تزايدت من حوالي 40 % من قيمة الشركات المتداولة بشكل عام في الولايات المتحدة الأمريكية في منتصف ثمانينيات القرن العشرين إلى حوالي ثلاثة أرباع قيمتها في بداية الألفية الجديدة. يعتبر هذا الاختلاف

ضخمًا للغاية. تشمل هذه الأصول غير الملموسة العلامة التجارية والموهبة وإستراتيجية العمل ــ أي شيء غير مادي ولا يعتبر جزءًا من نظام المحاسبة المالية الرسمي. ولكن توسع مفهوم الأصول غير الملموسة ليشمل البيانات التي تمتلكها وتستخدمها الشركة أيضًا.

في النهاية، ما يعنيه هذا الأمر أنه لا توجد طريقة واضحة لتقييم البيانات، حيث إنه في اليوم الذي بدأت فيه التعاملات بأسهم شركة فيس بوك، بلغ الاختلاف بين أصولها الرسمية وقيمتها غير الملموسة وغير المسجلة حوالي 100 مليار دولار. إنه لأمر غريب. ولكن يجب رأب فجوات مثل هذه وسيحدث من خلال إيجاد الشركات لطرق تسجل بواسطتها قيمة أصول بياناتها في تقارير موازنتها.

هناك بعض الخطوات الصغيرة التي تم اتخاذها في هذا الطريق، حيث قال أحد المديرين التنفيذيين لواحدة من كبريات شركات الشبكات المحمولة إن الشركة لاحظت القيمة الكبيرة للبيانات التي تملكها وتدرس حاليًّا ما إذا كان يمكن معاملتها على أنها من ضمن أصول المؤسسة طبقًا لأصول المحاسبة الرسمية. ولكن بمجرد أن سمع محامو الشركة بهذه المبادرة، أوقفوها على الفور، حيث إن وضع البيانات في الدفاتر المالية يجعل الشركة مسئولة قانونيًّا عنها، كما قال المستشارون القانونيون، الأمر الذي اعتقدوا أنه ليس من الأفكار الجيدة.

في الوقت ذاته، سيقوم المستثمرون هم أيضًا بالبدء في ملاحظة القيمة الاختيارية للبيانات، حيث إن أسعار أسهم الشركات التي تمتلك البيانات أو يمكنها الحصول عليها بسهولة سترتفع بشكل كبير، في حين سترى الشركات الأخرى التي لم تتمتع بالحظ ذاته أن قيمتها السوقية تنخفض. يجب ألا تظهر البيانات بشكل رسمي في تقارير الموازنة حتى يتحقق هذا الأمر، حيث ستقوم الأسواق والمستثمرون بتحديد أسعار لهذه الأصول غير الملموسة بتقديراتهم الخاصة ــ وإن كان بصعوبة، مثلما ظهر من تأرجح أسهم شركة فيس بوك في الأشهر القليلة الأولى من طرحها بالسوق. ولكن بعد أن هدأت الصعوبات المحاسبية والمخاوف القانونية، من المؤكد أن تظهر قيمة البيانات في تقارير الموازنة للمؤسسات على أنها فئة جديدة من الأصول المؤسسية.

كيف سيتم تقييم البيانات؟ إن حساب قيمتها لن يعني مجرد إضافة الأرباح التي نحصل عليها من استخدامها الأولي. فإن كانت أغلب قيمة البيانات ساكنة ويتم استخراجها من الاستخدامات الثانوية المستقبلية، فلن يمكن لأي شخص أن يقوم على الفور بتقديرها. يعتبر هذا الأمر مماثلاً لصعوبات تحديد أسعار المشتقات المالية قبل تطوير معادلة المدارس السوداء في سبعينيات القرن العشرين، أو صعوبة تحديد السواكن، في حين كانت المزادات العلنية والتبادلات والمبيعات الخاصة والتراخيص والكثير من النزاعات القضائية تخلق ببطء سوقًا من المعرفة. إذا لم يكن هناك مفر، فإن وضع سعر للقيمة الاختيارية للبيانات سيكون فرصة ذهبية للقطاع المالي.

من بين طرق البداية للقيام بهذا الأمر، فحص الإستراتيجيات التي يضعها مالكو البيانات لاستخراج القيمة منها. ستكون أكثر الاحتمالات وضوحًا لدى الشركات التي

تمتلك استخدامًا فعليًّا للبيانات. من غير المحتمل أن تتمكن أي من الشركات من الكشف عن جميع القيم الساكنة المخزنة في البيانات. من بين الأمور الأكثر طموحًا أن يتم إسناد استخراج القيم الساكنة من البيانات إلى أطراف أخرى. في عالم البيانات الضخمة، قد يرغب الكثير من مالكي البيانات في اختيار التوصل إلى اتفاقات للحصول على نسبة من القيم المستخرجة من البيانات بدلاً من الحصول على سعر ثابت لها، مثلما يفعل الناشرون عندما يدفعون نسبًا معينة من مبيعات الكتب أو الموسيقى أو الأفلام كأتعاب أدبية للمؤلفين أو الفنانين. كما أنها تشبه اتفاقات حقوق الملكية الفكرية في مجال التكنولوجيا الحيوية، حيث يطلب مانحو التراخيص الحصول على أتعاب أدبية على أية اختراعات تالية تنبثق من التكنولوجيا التي تم إعطاء الترخيص باستخدامها. بهذه الطريقة تمتلك الأطراف جميعها محفزات لتعظيم القيمة التي سيحصلون عليها من إعادة استخدام البيانات.

ولكن، ولأن الحاصل على ترخيص الاستخدام قد يفشل في استخراج القيمة الاختيارية للبيانات بشكل كامل، فقد لا يرغب مالكو البيانات في أن يتمكن من الوصول إلى جميع البيانات التي تمتلكها بشكل حصري، لذا يصبح "اختلاط البيانات" هو الأمر الشائع، وبهذه الطريقة يمكن لمالكي البيانات أن يوازنوا بين توقعاتهم للقيمة التي يحصلون عليها.

ظهر عدد من الأسواق لتجربة طرق تحديد أسعار للبيانات، من بينها سوق البيانات التي أنشئت في آيسلندا عام 2008، والتي تمنح إمكانية الوصول إلى مجموعات بيانية مجانية من مصادر بيانية أخرى، مثل الأمم المتحدة والبنك العالمي ويوروستات، وتحصل على الدخل من إعادة بيع البيانات التي يحصل عليها من موردي البيانات التجارية مثل شركات أبحاث السوق. حاولت شركات أخرى أن تلعب دور الوسيط المعلوماتي- أن تعمل كمنصات لأطراف أخرى تقوم من خلالها بعرض بياناتها سواء بالمجان أو بمقابل. تدور الفكرة حول أن يتمكن من يمتلك البيانات في قواعد بياناته من بيعها، كما يفعل موقع "إيباي" حيث يوفر للناس منصة لبيع الأشياء التي يحتفظون بها في عليِّة منازلهم. يشجع موقع Import.io الشركات على منح تراخيص استخدام لبياناتها والتي يمكن "جمع قصاصاتها" من خلال شبكة الإنترنت واستخدامها بالمجان. ولكن في الحقيقة، يقوم الموقع الذي أسسته موظفة جوجل السابقة "جيل إلباز"، بتوفير مجموعات بيانية تستغرق بعض الوقت حتى تتكدس.

دخلت شركة مايكروسوفت المعركة بسوق ويندوز أزور ماركت بلايس، الذي يهدف إلى التركيز على البيانات ذات الجودة العالية ويسعى لمعرفة ما يمكن الاستفادة به في العروض، كما تفعل شركة أبل من خلال إشرافها على العروض الموجودة في متجر تطبيقاتها. في رؤية شركة مايكروسوفت، قد يرغب مدير تسويق يعمل على برنامج إكسل في إعادة تصنيف البيانات الداخلية للشركة مقابل إجمالي نمو الناتج الداخلي الذي أعده المستشارون الاقتصاديون. ومن ثم يقوم بالضغط لشراء هذه البيانات التي تنتقل على الفور إلى أعمدة جداوله الموجودة على شاشة الحاسب.

حتى الوقت الحالي، لا توجد طريقة لإدراك مدى نجاح نماذج تقييم البيانات. ولكن الأمر المؤكد أن الاقتصاد قد بدأ في الاهتمام بالبيانات ــ ولكن يهتم الوافدون الجدد بالحصول على الفوائد، في حين قد تجد الشركات القديمة في هذا المجال استخدامات جديدة، كما يقول "تيم أورايلي"، الناشر التكنولوجي والعالم بوادي السيليكون: "البيانات هي القاعدة"، حيث إنها تعتبر حجر الأساس للبضائع والنماذج التجارية الجديدة.

إن النقطة الرئيسية فيما يتعلق بقيمة البيانات هي قدرتها اللامحدودة لإعادة استخدامها: قيمتها الاختيارية. إن جمع المعلومات أمر مهم ولكنه ليس كافيًا، حيث إن أغلب قيمتها تكمن في استخداماتها وليس في امتلاكها. في الفصل التالي، سنستعرض كيفية استخدام البيانات وشركات البيانات الضخمة التي تظهر.

7

ما تتضمنه البيانات

في عام 2011، فتحت إحدى الشركات الجديدة في سياتل تُدعى Decide.com أبوابها الإلكترونية بطموح جريء؛ حيث كانت ترغب في أن تكون محرك توقعات أسعار لعدد لا يُحصى من المنتجات الاستهلاكية. ولكن خططت الشركة لأن تبدأ بداية متواضعة إلى حدٍّ ما: توقع أسعار جميع الأجهزة الإلكترونية الممكنة من الهواتف المحمولة وشاشات التلفاز المسطحة وصولاً إلى الكاميرات الرقمية. قامت حواسب الشركة بمعالجة البيانات التي حصلت عليها من مواقع التجارة الإلكترونية ومشطت شبكة الإنترنت بحثًا عن أية معلومات أخرى عن أسعار المنتجات التي يمكنها الوصول إليها.

إن الأسعار على شبكة الإنترنت تتغير باستمرار حتى خلال اليوم الواحد؛ حيث تُحدث باستمرار من خلال عدد لا يُحصى من العوامل المعقدة، وبالتالي كانت الشركة بحاجة إلى جمع المعلومات طوال الوقت. ولم يكن الأمر مجرد بيانات ضخمة، بل كان "نصوصًا ضخمة" أيضًا؛ حيث إن النظام كان عليه أن يحلل الكلمات لإدراك ما إذا كان هناك منتج سيتوقف إنتاجه أو أن هناك نسخة جديدة منه بصدد طرحها في الأسواق، والمعلومات التي يجب أن يعرفها المستهلكون وتؤثر على الأسعار.

بعد عام واحد، كانت الشركة قادرة على تحليل 4 ملايين منتج باستخدام ما يزيد على 25 مليار ملاحظة تتعلق بالأسعار. كما تمكنت من تحديد الأمور الغريبة عن المتاجر والتي لم يكن بمقدور الناس معرفتها من قبل، مثل حقيقة أن أسعار الموديلات القديمة قد ترتفع بشكل مؤقت عند طرح موديلات جديدة، وقد يشتري أغلب الناس المنتجات القديمة اعتقادًا منهم أنها ستكون أرخص ثمنًا، ولكن بناءً على الوقت الذي يضغطون فيه على زر "شراء"، قد يدفعون أموالاً أكثر. وحيث إن المتاجر الإلكترونية أصبحت تستخدم بشكل متزايد أنظمة تحديد الأسعار الآلية، فإنه يمكن لموقع Decide.com أن يحدد ارتفاعات الأسعار الخوارزمية غير الطبيعية ويخبر المستهلكين بالانتظار. تبلغ دقة توقعات الشركة، طبقًا لقياساتها الداخلية، نسبة 77 % في وقت الشراء وتمكن المشترون من توفير حوالي 100 دولار في المتوسط لكل منتج يشترونه. ومن شدة ثقة الشركة بتوقعاتها, فإنها قررت أنه في حالة خابت توقعاتها ستقوم بدفع فارق الأسعار لمستخدمي الخدمة.

للوهلة الأولى، يبدو أن شركة Decide.com مثلها مثل الكثير من الشركات الجديدة الواعدة تهدف إلى توظيف المعلومات بطرق جديدة تجني من ورائها أرباحًا مقابل الجهود التي قامت بها. إن ما يجعل هذه الشركة مميزة ليس البيانات: تعتمد الشركة على المعلومات التي أخذت حق استخدامها من مواقع التجارة الإلكترونية والمعلومات المتفرقة من شبكة الإنترنت المجانية. كما أن الشركة ليست

متخصصة في الأمور التقنية: لم تفعل الشركة أيًّا من الأمور المعقدة للغاية التي يقوم بها بعض من المهندسين المتميزين في العالم بل اعتمدت على المهندسين الذين يعملون فيها. ولكن رغم أن جمع البيانات والمهارات التقنية من الأمور المهمة، فإن جوهر تميز الشركة يكمن في الفكرة: أي أن الشركة تمتلك "عقلية بيانات ضخمة". لقد لاحت للشركة في الأفق فرصة وأدركت أن بعض البيانات يمكن التنقيب فيها للكشف عن الأسرار القيمة الكامنة فيها. وإن كان هناك تشابه بين شركة Decide.com وشركة توقع تذاكر الطيران فايركاست، فسيكون هناك تفسير جيد لهذا الأمر: إن كلتيهما وليدة أفكار "أورين إتزيوني".

ناقشنا في الفصل السابق أن البيانات أصبحت مصدرًا جديدًا للقيمة المادية بفضل ما أطلقنا عليه القيمة الاختيارية لها، عند استخدامها في استخدامات مبتكرة. كان تركيزنا منصبًّا على الشركات القادرة على جمع البيانات، أما الآن فسيتحول تركيزنا لينصبّ على الشركات التي تستخدم البيانات، وكيف يمكنها أن تحتل موقعًا جيدًا في حلقة القيم المعلوماتية. وسوف نضع في اعتبارنا ما يعنيه هذا الأمر بالنسبة للمؤسسات وللأفراد، سواء في حياتهم العملية أو الشخصية.

هناك ثلاثة أنواع من شركات البيانات الضخمة قد ظهرت فجأة، والتي يمكن تصنيفها على أساس القيمة التي تقدمها. فكر فيها على أنها البيانات والمهارات والأفكار.

النوع الأول هو البيانات. هذه هي الشركات التي تمتلك البيانات أو يمكنها الحصول عليها على أقل تقدير. ولكن ربما لم يكن هذا هو السبب الذي دخلت من أجله هذه الشركات ذلك المجال: أو لا تمتلك المهارات المناسبة لاستخراج القيمة منها أو الخروج بأفكار مبتكرة عن القيم التي يمكن إطلاقها. أفضل مثال على ذلك هو موقع تويتر، الذي من الجلي أنه يتمتع بتدفقات هائلة من البيانات تمر خلال خوادمه ولكنه تحول إلى شركتين أخريين مستقلتين ليعطيهما حق استغلال تلك البيانات.

النوع الثاني هو المهارات، والتي عادة ما تكون الشركات الاستشارية والتكنولوجية والقائمة بالتحليلات والتي تمتلك خبرات خاصة وتستطيع القيام بالعمل، ولكن من المحتمل ألا تكون من مالكي البيانات ولا تمتلك القدرة على ابتكار استخدامات مبتكرة لها. مثل حالة وولمارت وبوب تارتس على سبيل المثال؛ حيث تحول التاجر إلى المتخصصين في شركة تيراداتا، إحدى شركات التحليلات، للمساعدة على ابتكار الأفكار.

النوع الثالث هو عقلية البيانات الضخمة. بالنسبة لبعض الشركات، لم تكن البيانات ولا كيفية استخدامها هي المقومات الأساسية لنجاحها. بل إن ما جعلها متميزة هو أن مؤسسيها وموظفيها يملكون أفكارًا متميزة عن طرق استغلال البيانات لاستخراج أنواع جديدة من القيم الكامنة بها. من بين الأمثلة على هذا النوع "بيتي واردن"، مهووسة الحواسب التي شاركت في تأسيس موقع "جيتباك"، الذي يقدم توصيات للقيام بالرحلات بناء على الصور التي يرفعها المستخدمون على خوادم الموقع.

حتى الوقت الحالي يلاقي النوعان الأولان من الشركات أكبر قدر من الاهتمام:

المهارات التي أصبحت نادرة هذه الأيام والبيانات، التي تتوافر بقدر كبير. ظهرت مهنة جديدة في الأيام الأخيرة وهي "مهندس البيانات"، والتي تتطلب مهارات الإحصائيين ومبرمجي برامج الحاسب ومصممي البيانات التصويرية التفاعلية ورواة القصص مجتمعة. بدلاً من إجهاد أعينهم خلف عدسات المجهر لحل ألغاز الكون، يمشط مهندسو البيانات المجموعات البيانية لاكتشاف ما تحتوي عليه. قدم معهد ماكينزي الدولي توقعات أليمة فيما يتعلق بندرة وجود مهندسي البيانات في الوقت الحالي وفي المستقبل (الأمر الذي يستشهد به مهندسو البيانات في الوقت الحالي ليشعروا بأنهم متميزون ومن أجل أن يرفعوا من رواتبهم).

عادة ما كان يطلق "هال فاريان"، كبير علماء الاقتصاد بشركة جوجل، على مهنة الإحصائيين أنها "أروع" مهنة في العالم؛ حيث يقول: "إذا ما أردت أن تكون ناجحًا، فسترغب في أن تكون من يتمم الأمور وأن تكون من القلة القليلة التي تفهم في أمر منتشر بشكل كبير ورخيص الثمن. إن البيانات متوافرة بشكل كبير وعلى قدر كبير من الأهمية ولكن الأمر النادر هو معرفة كيفية استخراج المعرفة منها. لهذا السبب سيصبح الإحصائيون ومديرو قواعد البيانات والمتخصصون في التعليم الإلكتروني في موقع متميز للغاية".

ولكن يبدو أن كل هذا الاهتمام بالمهارات والتقليل من أهمية البيانات سيثبت أنه لن يستمر طويلاً؛ حيث إنه مع تطور المجال سيتم التغلب على ندرة من يفهمون به وإن المهارات التي تباهي بها "فاريان" ستصبح عادية. هذا إلى جانب الاعتقاد الخاطئ بأنه بفضل وجود كم كبير متوافر من البيانات فإنه من الطبيعي الحصول عليها بالمجان أو أن قيمتها بخسة. في حقيقة الأمر، إن البيانات هي المكون الرئيسي. ولكي تعرف السبب، تأمل في المكونات المختلفة لسلسلة قيمة البيانات، وكيف أنه من المحتمل أن تتغير بمرور الوقت. يمكننا البدء باستعراض كل مكون على حدة ـــ مالكي البيانات ومتخصصي البيانات وعقلية البيانات الضخمة.

سلسلة قيمة البيانات الضخمة

إن المكون الرئيسي للبيانات الضخمة هو المعلومات ذاتها؛ لذا من المنطقي أن نتناول مالكي البيانات أولاً. قد لا يكونون هم من قاموا بجمع البيانات بأنفسهم، ولكنهم يملكون القدرة على الوصول إلى المعلومات ليستخدموها بأنفسهم أو يكلفوا أطرافًا أخرى باستخراج القيمة منها. على سبيل المثال، برنامج ITA، شبكة حجز تذاكر رحلات الطيران العملاقة (التي يأتي ترتيبها بعد كل من أماديوس وترافلبورت وسابري)، قام بإمداد موقع فايركاست بالبيانات من أجل توقعاته بأسعار تذاكر رحلات الطيران، ولكنه لا يقوم بالتحليل بنفسه. لماذا لا؟ لقد اعتقدت شبكة ITA أن عملها يقوم على استخدام البيانات للغرض الذي تم تصميمها من أجله ـــ بيع تذاكر رحلات الطيران ـــ وليس من أجل بعض الاستخدامات الإضافية. ولهذا السبب كانت كفاءتها الجوهرية متميزة إلى جانب أنها كانت بحاجة إلى الاستعانة بابتكارات "إتزيوني".

علاوة على ذلك، اختارت الشركة ألا تستخدم البيانات بنفسها بسبب الموقع الذي

تحتله في سلسلة قيمة المعلومات. يقول "كارل د. يماركن"، أحد مؤسسي الشركة ورئيس قسم التكنولوجيا السابق بها: "تجنبت الشركة العمل في المشروعات التي تتطلب الاستخدام التجاري المباشر للبيانات فيما يتعلق بأرباح تذاكر خطوط الطيران. إن الشركة تمتلك قدرة خاصة على الوصول إلى تلك البيانات التي تحتاج إليها الشركة من أجل تحسين خدماتها ولم تحاول تعريض كل هذا للخطر". وبدلاً من هذا تحولت إلى إسناد حق استخدام البيانات إلى أطراف أخرى بدلاً من استخدامها بنفسها. ذهبت أغلب القيم الثانوية لهذه البيانات إلى شركة فايركاست: إلى عملائها في شكل تذاكر أرخص سعرًا وإلى موظفيها ومالكيها من الدخل الذي حصلت عليه الشركة من الإعلانات والعمولات وفي النهاية من بيع الشركة.

قامت بعض الشركات بتحديد موقعها بمهارة في منتصف تدفقات المعلومات بحيث يمكنها أن تحصل على كم كبير من البيانات وتحصل على القيم منها. كان هذا ما قامت به الشركات العاملة في مجال البطاقات الائتمانية في الولايات المتحدة الأمريكية. فطوال سنوات، قادت تكلفة مقاومة التزوير العالية الكثير من البنوك الصغيرة ومتوسطة الحجم إلى تجنب إصدار بطاقاتها الائتمانية وأن تسند عمليات البطاقات الائتمانية إلى المؤسسات المالية الكبرى، والتي تمتلك الحجم والقدرة على الاستثمار في هذه التكنولوجيا؛ حيث تسببت الموازنات العامة لبعض الشركات على غرار كابيتال وان وبنك أمريكا في زيادة الاهتمام بهذا المجال. ولكن ندمت البنوك الأصغر حجمًا في الوقت الحالي على هذا القرار؛ حيث إن التخلي عن عمليات بطاقات الائتمان قد تسبب في حرمانها من الحصول على البيانات عن أنماط الإنفاق والتي كانت ستسمح لها بمعرفة المزيد عن عملائها بحيث تتمكن فيما بعد من أن تبيع لهم خدمات خاصة بهم.

ولكن بدلاً من هذا بدا أن البنوك الكبرى والشركات المصدرة للبطاقات الائتمانية مثل فيزا وماستر كارد قد احتلت موقعًا متميزًا من سلسلة القيمة المعلوماتية، فمن خلال تقديم خدماتها للكثير من البنوك والشركات التجارية، تمكنت من تعقب المزيد من التعاملات التي كانت تُجرى على شبكاتها واستخدمتها في استنتاج الكثير عن سلوكيات المستهلكين: تحولت نماذج عملها من مجرد تحليل المدفوعات إلى جمع البيانات. السؤال الذي طرأ حينها هو: ما الذي ستفعله بهذه البيانات؟

كان بمقدور شركة ماستر كارد إسناد البيانات التي جمعتها إلى طرف ثالث الذي قد يتمكن من استخراج القيمة منها، كما فعلت شركة ITA، ولكن فضلت أن تقوم بتحليل البيانات بنفسها؛ حيث قامت بإنشاء قسم أطلقت عليه مرشدي ماستر كارد والذي تمكن من تكديس وتحليل 65 مليار تعامل من 1.5 مليار من حاملي البطاقات الائتمانية من 210 دول من أجل تحسين التوجهات العملية وتوجهات المستهلكين. ومن ثم قامت ببيع هذه المعلومات إلى آخرين؛ حيث اكتشفت من بين أمور أخرى أنه إن قام الناس بملء خزانات وقود سياراتهم في حوالي الساعة الرابعة عصرًا، فمن المرجح أن ينفقوا ما بين 35 و50 دولارًا في الساعة التالية في متجر للمواد الغذائية أو في أحد المطاعم؛ حيث يمكن للمسوقين أن يستخدموا تلك المعلومات في طباعة قسائم لأحد المتاجر القريبة على الوجه الخلفي لإيصالات محطة الوقود

لاستخدامها في هذا الوقت من اليوم.

من خلال عملها كوسيط يقف في منتصف تدفقات المعلومات، أصبحت شركة ماستر كارد في موقع متميز لجمع المعلومات واستخراج القيم منها. يمكن للمرء أن يتوقع في المستقبل أن تتخلى شركات البطاقات الائتمانية عن العمولات على التعاملات؛ حيث ستقوم بها بالمجان في مقابل الحصول على المزيد من البيانات، ومن ثم تحصل على أرباح مقابل بيعها تحليلات دقيقة بناءً على تلك البيانات.

المكون الثاني هو متخصصو البيانات: الشركات التي تملك الخبرة أو التكنولوجيا اللازمة للقيام بالتحليلات المعقدة. اختارت شركة ماستر كارد أن تقوم بهذا الأمر داخليًّا، في حين كانت بعض الشركات تتنقل بين الفئات، وتحول الكثير منها إلى المتخصصين. على سبيل المثال، تعمل شركة الاستشارات "أسينتشر" مع شركات من مختلف الصناعات من أجل تركيب حساسات لا سلكية متطورة والقيام بتحليل البيانات التي تجمعها تلك الحساسات. في مشروع رائد مع مدينة سانت لويس بولاية ميزوري، قامت الشركة بتركيب حساسات في عدد من حافلات النقل العام من أجل مراقبة حالة محركاتها لتوقع الأعطال أو لتحديد الوقت المتوقع للقيام بالصيانة الدورية؛ مما أدى إلى التقليل من تكلفة صيانتها بنسبة تصل إلى 10 %. وكان هناك اكتشاف جيد ـــ أن المدينة يمكن أن تؤجل تغيير قطع الغيار المخطط لتغييرها كل 200 ألف ـــ250 ألف ميل لتغيرها كل 280 ألف ميل؛ مما وفر ما يزيد على ألف دولار لكل مركبة. وجني العميل، وليس الشركة الاستشارية، قيمة البيانات.

في عالم البيانات الطبية، نجد مثالاً رائعًا على كيف يمكن لشركات التكنولوجيا الخارجية أن تقدم خدمات مفيدة. مركز مستشفى ميدستار واشنطن بالعاصمة الأمريكية واشنطن، والذي يعمل بالتعاون مع مركز أبحاث شركة مايكروسوفت ويستخدم برنامج أمالجا المقدم من شركة مايكروسوفت، والذي قام بتحليل أعوام من سجلاته الطبية المجهَّلة ـــ البيانات الديموغرافية للمرضي وفحوصاتهم وتشخيصاتهم والعلاج الذي تلقوه وغيرها الكثير ـــ من أجل تقليل معدلات الانتكاسات والعدوى. تعتبر هذه البيانات من أعلى جوانب الرعاية تكلفة، الأمر الذي يعني أن القيام بأي شيء من أجل تخفيض هذه المعدلات سيوفر الكثير على الناس.

كشف هذا الأسلوب عن بعض العلاقات التبادلية المدهشة. من بين النتائج القائمة بجميع الظروف التي أدت لزيادة احتمالات عودة المرضى الذين تم شفاؤهم للمستشفى مرة أخرى خلال شهر. بعض هذه الظروف معروفة ولا يوجد لها حل سهل. المرضى الذين يعانون فشلاً احتقانيًّا بالقلب من المرجح أن يعودوا للمستشفى مرة أخرى: حيث إنها حالة مرضية من الصعب علاجها. ولكن تمكن النظام أيضًا من اكتشاف بعض التوقعات المفاجئة: الحالة العقلية للمرضى. تزداد احتمالية عودة المرضى للمستشفى في خلال شهر من صرفهم من المستشفى بشكل ملحوظ إذا ما كانت شكواهم المبدئية تحتوي على كلمات تتعلق بحالة

عقلية، مثل "الاكتئاب".

رغم أن هذه العلاقات التبادلية لا تقول أي شيء لإثبات الأسباب، فإنها اقترحت أن التدخل الطبي لعلاج الحالة العقلية للمريض قبل صرف المريض من المستشفى قد يتسبب في تحسين حالتهم البدنية هي الأخرى؛ مما يقلل من معدلات عودة المرضى للمستشفى ويقلل من تكاليف الرعاية الطبية. هذه النتائج التي اكتشفتها أجهزة الحاسب داخل كميات هائلة من البيانات تعتبر من الأمور التي لم يكن بمقدور الأشخاص الدارسين للبيانات أن يكتشفوها مهما طال بهم الزمن. لم تكن شركة مايكروسوفت من المسيطرين على البيانات التي كانت مملوكة للمستشفى، ولم تكن تملك فكرة مذهلة، الأمر الذي لم يكن مطلوبًا في هذه الحالة، ولكنها قامت بدلاً من كل ذلك بتقديم الأداة البرمجية اللازمة، برنامج أمالجا، لاكتشاف الأفكار.

تعتمد الشركات التي تمتلك البيانات الضخمة على المتخصصين لاستخراج القيمة من البيانات، ولكن رغم الإطراء الكثير والمسميات الوظيفية الأنيقة مثل "نينجا البيانات"، فإن حياة الخبراء التقنيين لا تكون مبهجة دائمًا مثلما قد تبدو؛ حيث إنهم يكدون في التنقيب داخل مناجم قيم البيانات ويحصلون على رواتب عالية، إلا أنهم يسلمون هذه الكنوز التي اكتشفوها للملاك الحقيقيين للبيانات.

المكون الثالث يتكون من الشركات والأفراد الذين يمتلكون عقلية البيانات الضخمة الذين تكمن قوتهم في رؤيتهم الفرص قبل غيرهم ــ حتى إن كانوا يفتقدون البيانات أو المهارات ليغتنموا هذه الفرص. وربما كان عدم امتلاكهم هذه الأمور، كوافدين جدد على المجال، هو ما جعل عقولهم متحررة من القيود الخيالية: إنهم يرون الأمور التي يمكن تحقيقها بدلاً من أن يتقيدوا بالأمور التي يمكن تنفيذها.

جسد "برادفورد كروس" ما تعنيه عبارة عقلية البيانات الضخمة. في أغسطس عام 2009 عندما كان في منتصف العشرينات من عمره، قام ومجموعة من أصدقائه بإنشاء موقع FlightCaster.com. كان الموقع يقوم بنفس وظيفة موقع F؛ حيث كان يقدم توقعات عن احتمالات تأخر الرحلات الجوية في الولايات المتحدة الأمريكية. من أجل تقديم هذه التوقعات بشكل دقيق، قام الموقع بتحليل بيانات جميع الرحلات الجوية للسنوات العشر الماضية، مقارنة ببيانات الطقس الماضية والحالية.

الأمر المذهل أن مالكي البيانات أنفسهم لم يتمكنوا من القيام بهذا الأمر؛ حيث لم يكن أي منهم يمتلك الحافز ــ أو الحاجة الملحة ــ لاستخدام البيانات بهذه الطريقة. في حقيقة الأمر، إذا ما تجرأت مصادر بيانات مكتب النقل الأمريكي وإدارة الطيران الفيدرالية والإدارة الوطنية لدراسة المحيطات والغلاف الجوي وخدمة الطقس الوطنية ــ إذا ما تجرأت على توقع تأخر الرحلات الجوية التجارية، فسيكون مجلس النواب الأمريكي قد عقد جلسات استماع وسيبدأ رؤساء المكاتب الخدمية في التشاجر. ولم تتمكن شركات الخطوط الجوية من فعل ذلك هي الأخرى ــ أو لم ترغب في هذا، حيث إنها تستفيد من خلال الحفاظ على أدائها

كوسيط خفيّ قدر الإمكان. ولكن، تطلب تحقيق هذا الأمر عددًا من المهندسين المختفين. في حقيقة الأمر، أصبحت توقعات موقع FlightCaster دقيقة للغاية لدرجة أن موظفي شركات الخطوط الجوية قد بدءوا في استخدامها: لا ترغب شركات الخطوط الجوية في الإعلان عن تأخر الرحلات حتى الدقيقة الأخيرة؛ لذا فإنه رغم أنها هي المصدر الرئيسي للمعلومات، فإنها ليست أسرع مصدر لهذه المعلومات.

بفضل عقلية البيانات الضخمة التي تمتلكها ــ إدراكها لإمكانية معالجة البيانات المتواجدة على شبكة الإنترنت بشكل علني وبطريقة تسمح بتقديم أجوبة قد يحتاج إليها الملايين من البشر ــ كان موقع FlightCaster الذي أنشأه "كروس" هو الموقع الأول في هذا المجال، ولكن بالكاد، ففي الشهر نفسه الذي تم إطلاق الموقع فيه، بدأ العباقرة الذين أنشأوا موقع FlyOnTime.us في جمع البيانات مفتوحة المصدر من أجل بناء موقعهم. في عام 2011 قام "كروس" وشركاؤه ببيع شركتهم إلى "نيكست جامب"، الشركة التي تدير برامج تخفيضات المؤسسات باستخدام تقنيات البيانات الضخمة.

بعد ذلك تحول "كروس" إلى مجال قديم آخر رأى فيه ثغرة يمكن للوافد الجديد أن يدخل من خلالها: الإعلام الإخباري. كانت شركته الجديدة التي أطلق عليها اسم "بريسماتيك"، تقوم بتكديس محتويات شبكة الإنترنت وتصنيفها على أساس تحليل النصوص ومفضلات المستخدمين والموضوعات المتعلقة بشعبية مواقع التواصل الاجتماعي وتحليل البيانات الضخمة. الأمر المهم، أن النظام لم يكن يفرق بشكل كبير بين الموضوعات المنشورة على مدونة للمراهقين أو موقع مؤسسة ما ومقال منشور بجريدة واشنطن بوست: إذا ما كان المحتوى يعتبر ذا صلة بالموضوع ويحظى بشعبية جيدة (من خلال عدد مرات تصفحه وعدد مرات إعادة نشره)، فسيظهر في أعلى الصفحة.

تعتبر شركة بريسماتيك، كخدمة، إدراكًا لطرق تفاعل الجيل الجديد مع وسائل الإعلام؛ حيث إنهم أصبحوا يعتقدون أن مصادر المعلومات قد فقدت أهميتها الأساسية. يعتبر هذا الأمر تذكرة متواضعة لأباطرة وسائل الإعلام السائدة بأن المجتمع قد أصبح أكثر إلمامًا بمجريات الأمور أكثر مما يفعلون هم، وأصبح من المحتم على هؤلاء الصحفيين المقيدين أن ينافسوا المدونين في مجالهم. ولكن النقطة الرئيسية في هذا الأمر أنه من الصعب التخيل بأن بريسماتيك قد ظهرت من داخل مجال الإعلام، رغم حقيقة أنها قادرة على جمع كم كبير من المعلومات. لم تفكر الشركات الأعضاء في نادي الإعلام الوطني أبدًا في إعادة استخدام البيانات من خلال شبكة الإنترنت المتعلقة بالاستهلاك الإعلامي. ولم يفكر خبراء تحليل البيانات في آرمونك بنيويورك أو بنجالور بالهند، في توظيف المعلومات بهذه الطريقة. لقد تطلب الأمر وجود "كروس"، الوافد الجديد ذي الشعر الأشعث والمتكاسل، ويمكنه لاستكمال هذا الأمر من خلال استخدام البيانات الضخمة أن يخبر العالم بأنه عليه الانتباه إلى أمور أفضل من محرري جريدة نيويورك تايمز.

إن فكرة عقلية البيانات الضخمة، والدور الذي لعبه أحد الوافدين الجدد المبتكرين الذي امتلك فكرة رائعة، لا يشبهان ما حدث في بدايات التجارة الإلكترونية في

منتصف تسعينيات القرن العشرين، عندما لم يتقيد رواد هذا المجال بالأفكار التقليدية أو القيود المؤسسية للصناعات الأقدم؛ لذا قامت شراكة متساوية الأسهم، ليست بارنز ونوبل، بإنشاء مكتبة إلكترونية (أمازون المقدمة بواسطة جيف بيزوس). قام أحد مطوري البرمجيات، وليس "سوثيبي"، ببناء موقع للمزادات العلنية (إباي المقدم بواسطة بيير أوميديار). في الوقت الحالي، عادة ما لا يمتلك رواد عقليات البيانات الضخمة البيانات في البداية، ولكن لهذا السبب، لا يمتلكون أيضًا الاهتمامات الملحة والعوائق المالية التي قد تمنعهم عن إطلاق العنان لأفكارهم.

كما رأينا من قبل، هناك حالات تقوم فيها الشركات بالمزج بين الكثير من سمات البيانات الضخمة تلك. ربما امتلك "إتزيوني" و"كروس" الأفكار قبل أن تراود الآخرين، ولكنهم كانوا يملكون المهارة أيضًا. إن الأيدي العاملة في مصانع تيرداتا وأسينتور لم تكن تثبت حضورها للعمل فقط ولكن كان المعروف عنهم أنهم يخرجون بأفكار رائعة من وقت لآخر. ولكن النماذج الأولية تعمل كطرق لتقدير الأدوار التي لعبتها الشركات المختلفة. يأتي رواد البيانات الضخمة في الوقت الحالي من خلفيات مختلفة وينقلون تطبيق مهاراتهم البيانية على مجموعة متنوعة من المجالات. هناك جيل جديد يظهر من الممولين ورواد الصناعات الجديدة، وعادة ما يكونون من العاملين السابقين بشركة جوجل أو ممن يطلق عليهم اسم مافيا باي بال (قادة الشركة السابقون مثل بيتر ثييل وريد هوفمان وماكس ليفشن)، والذين يعتبرون، إلى جانب عدد قليل من مهندسي علوم الحاسب الأكاديميين، كبار مؤيدي الشركات الجديدة العاملة في مجال البيانات في الوقت الحالي.

إن الرؤية المبتكرة للأفراد والشركات في السلسلة الغذائية للبيانات الضخمة تساعدنا على إعادة تقييم قيمة الشركات. على سبيل المثال، ربما لم يكن موقع Sa المنصة المفيدة للشركات لاستضافة تطبيقاتها المؤسسية فحسب: بل يعتبر أيضًا ذا موقع متميز فيما يتعلق بإطلاق القيمة من البيانات التي تتدفق على بنيته التحتية. تجمع شركات شبكات الهواتف المحمولة، كما رأينا في الفصل السابق، عددًا هائلاً من البيانات ولكنها عادة ما لا تتمكن من رؤية القيمة المختزنة بها. يمكن لتلك الشركات أن تسند استخدام البيانات إلى أطراف أخرى يمكنها استخراج قيم جديدة من البيانات ــ كما فعلت تويتر عندما قررت إسناد حقوق استخدام بياناتها إلى شركتين أخريين.

قامت بعض المؤسسات سعيدة الحظ بالفصل بين نطاقات العمل كنوع من الوعي الإستراتيجي. شركة جوجل التي تقوم بجمع البيانات مثل عبارات استفسارات البحث، واتتها فكرة رائعة باستخدام تلك البيانات في إنشاء برنامج للتدقيق الإملائي، وكانت تمتلك في داخلها المهارات اللازمة لإتمام هذا البرنامج على الوجه الأمثل. مع وجود الكثير من الأنشطة الأخرى، استفادت شركة جوجل من الاندماج العمودي بسلسلة قيمة البيانات الضخمة؛ حيث إنها تحتل جميع المواقع الثلاثة في الوقت ذاته. في الوقت نفسه، توفر شركة جوجل بعضًا من بياناتها إلى

الآخرين من خلال واجهة برمجة التطبيقات بحيث يمكن إعادة استخدامها ويتم الحصول على قيمة إضافية منها. من بين الأمثلة على هذا الأمر، خرائط جوجل، التي يمكن للجميع استخدامها على شبكة الإنترنت بدءًا من شركات العقارات إلى المواقع الإلكترونية الحكومية مجانًا (رغم أن المواقع التي بها عدد كبير من المستخدمين يكون لزامًا عليها دفع مقابل استخدامها).

شركة أمازون هي الأخرى تمتلك العقلية والخبرة والبيانات. في حقيقة الأمر، بدأت الشركة نموذج عملها على هذا الأساس، الأمر الذي يعتبر منافيًا للمعتاد. ففي البداية واتتها فكرة نظام التوصيات وحده. شرحت دعوتها للاكتتاب العام في سوق الأوراق المالية عام 1997 مبدأ "الترشيح التعاوني" قبل حتى أن تدرك الشركة كيف سيعمل أو كانت تمتلك البيانات الكافية للاستفادة منه.

لقد تجاوزت كل من جوجل وأمازون جميع التصنيفات، ولكن من خلال إستراتيجيات مختلفة. عندما اهتمت جوجل في المقام الأول بجمع أي نوع من البيانات، كانت تضع في اعتبارها الاستخدامات الثانوية لها. جمعت سيارات استكشاف الشوارع، كما رأينا من قبل، معلومات نظام تحديد المواقع العالمي ليس فقط من أجل خدمة الخرائط التي تقدمها ولكن أيضًا من أجل تدريب السيارات التي تقود نفسها. وعلى النقيض، تركز شركة أمازون على الاستخدامات الرئيسية وتستخدمها في استخدامات فرعية فقط كنوع من الربح الهامشي. يعتمد نظام التوصيات الخاص بالشركة، على سبيل المثال، على بيانات ضغطات المستخدمين كمؤشر، ولكن لم تستخدم الشركة هذه المعلومات للقيام بأمر خارق للعادة مثل توقع حالة الاقتصاد القومي أو أوبئة الأنفلونزا.

رغم أن قراء برنامج كيندل المقدم من أمازون يمكّنهم أن يعرفوا ما إذا كانت إحدى صفحات الكتاب قد تم التعليق عليها أو التأكيد عليها من قبل الكثير من المستخدمين، فإن الشركة لا تقوم ببيع هذه المعلومات إلى الناشرين أو المؤلفين. قد يتوق المسوقون إلى معرفة أي من الفقرات تتمتع بشعبية كبيرة ويستخدمون هذه المعلومات من أجل تحسين عملية بيع الكتب. قد يرغب المؤلفون في معرفة في أية مرحلة يتوقف القراء عن قراءة مجلداتهم الكبيرة، ليستخدموا هذه المعلومات في تحسين أعمالهم. قد يلاحظ الناشرون الموضوعات التي ستكون موضوع الكتاب الكبير القادم. ولكن يبدو أن أمازون تميل إلى ترك مجال البيانات هادئًا كما هو.

عند توظيفها بمهارة، يمكن للبيانات الضخمة أن تغير من نماذج العمل للشركات وكذلك طرق التعامل بين الشركاء التي لم تتغير منذ زمن طويل. في إحدى الحالات المدهشة، قامت إحدى شركات السيارات الأوربية الكبرى بإعادة تشكيل علاقتها التجارية مع أحد موردي قطع الغيار من خلال توظيف استخدام البيانات المتعلقة بقطع الغيار التي تفتقدها الشركة صانعة السيارات. (ولأننا قد حصلنا على هذا المثال على أساس واحدة من الشركات الأساسية التي ساعدت على تحليل البيانات، فإننا للأسف لن نستطيع إفشاء أسماء هذه الشركات).

تمتلئ السيارات هذه الأيام بالكثير من الرقاقات الإلكترونية والحساسات

والبرامج التي تنقل بيانات الأداء إلى حواسب مصنعيها عند خضوعها للصيانة. أصبحت السيارات من الفئة المتوسطة في الوقت الحالي تحتوي على حوالي 40 من المعالجات الدقيقة؛ حيث إن جميع الأجهزة الإلكترونية التي تحتوي عليها السيارة تمثل حوالي ثلث تكلفتها. الأمر الذي يجعل السيارات في الوقت الحالي تشبه ما أسماه "موراي" في السفن "مراصد الإبحار". إن القدرة على جمع البيانات عن كيفية عمل قطع غيار السيارات في أثناء عملها ــ وإعادة استخدام هذه البيانات من أجل تحسينها ــ تبين أنها ميزة تنافسية مهمة بالنسبة للشركات التي يمكنها الحصول على هذه المعلومات.

بالعمل مع شركات خارجية متخصصة في تحليل البيانات، جعلت صانع السيارات يكتشف أن أحد الحساسات في خزان الوقود والذي صنعه أحد الموردين الألمان يعمل بشكل سيئ؛ حيث كان يصدر الكثير من التحذيرات الزائفة مقابل تحذير صحيح واحد. وكان بمقدور الشركة أن تسلم هذه المعلومات إلى المورد وطلبت منه إصلاح الأمر. ربما فعلت هذا الأمر فقط بفضل عصر الأعمال الأكثر احترامًا الذي نعيشه حاليًّا، ولكن الشركة صانعة السيارات كانت تنفق أموالاً طائلة على برنامجها لتحليل البيانات. ورغبت في استخدام هذه المعلومات من أجل استرداد بعض من استثماراتها.

تأملت الشركة في الخيارات المتاحة لها. هل عليها أن تبيع البيانات؟ كم ستكون قيمة المعلومات؟ ماذا لو فشل المورد في إصلاح الأخطاء وأصبحت الشركة ملزمة بالعمل بقطعة غيار لا تعمل بشكل جيد؟ أدركت الشركة أنها إن سلمت هذه المعلومات إلى المورد فإن قطع الغيار المماثلة التي يتم تركيبها في سيارات المنافسين ستُحسن هي الأخرى. أما التأكد من أن سياراتها فقط هي التي ستستفيد من هذا التحسين كان من الخطوات التي تدل على الفطنة الشديدة. في نهاية الأمر، راودت الشركة المصنعة للسيارات فكرة مبتكرة؛ فقد وجدت طريقة لتحسين قطعة الغيار باستخدام برنامج إلكتروني معدل؛ حيث حصلت على القيمة الكامنة في هذه البيانات ومن ثم باعتها إلى المورد نفسه ــ وحصلت على ربح جيد مقابل هذه العملية.

وسطاء البيانات الجدد

من الذي يمتلك أعلى قيمة من البيانات في سلسلة قيم البيانات الضخمة؟ في الوقت الحالي، قد تكون الإجابة هي أولئك الذين يمتلكون عقلية البيانات الضخمة، والأفكار المبتكرة. كما شاهدنا في عصر الإنترنت، إن الشركات التي تمتلك القدرة على التحرك أولاً هي التي يمكنها النجاح. ولكن قد لا تدوم الميزة طويلاً؛ حيث إنه بتقدم عصر البيانات الضخمة للأمام، قد تعتنق شركات أخرى عقلية البيانات الضخمة وستنتهي أفضلية الرواد الأوائل.

هل يمكن حينها إذن أن يتحول أساس قيمة البيانات ليكون المهارات؟ رغم كل شيء، إن منجم الذهب لا يساوي شيئًا إذا لم تتمكن من استخراج الذهب منه. ولكن يفترض عصر الحوسبة عكس ذلك. في الوقت الحالي أصبحت الوظائف التي تعمل

في مجالات إدارة قواعد البيانات وعلوم البيانات والتحليل البياني ومعادلات التعليم الإلكتروني والمهن المشابهة مطلوبة للغاية. ولكن بمرور الوقت ستصبح البيانات الضخمة جزءًا عاديًّا من الحياة اليومية، ومع تحسن الأدوات وسهولة استخدامها، ومع حصول الكثير من الأشخاص على الخبرة اللازمة في هذا المجال، فإن قيمة المهارات هي الأخرى ستقل لهذا السبب. بالمثل، أصبحت قدرات برمجة الحواسب معتادة في الفترة ما بين منتصف الستينيات والثمانينيات من القرن العشرين. في الوقت الحالي، قللت شركات التعهيد الخارجي من قيمة القدرات البرمجية أكثر من ذي قبل؛ حيث إن ما كان يعتبر مثالاً نموذجيًّا على الفطنة التقنية أصبح في الوقت الحالي محركًا لتطوير فقراء العالم. ولكن هذا لا يعني أن خبرات البيانات الضخمة غير ذات أهمية، بل إنها ليست المصدر الأكثر أهمية للقيمة؛ حيث إن المرء يمكنه الحصول عليها من الخارج.

في الوقت الحالي، في المراحل الأولى من البيانات الضخمة، يبدو أن الأفكار والمهارات هي التي تمتلك القيمة الكبرى. ولكن في النهاية ستظل أغلب القيمة في البيانات نفسها، هذا لأننا سنكون أكثر قدرة على فعل المزيد باستخدام المعلومات وكذلك لأن مالكي البيانات سيصبحون أكثر قدرة على تقدير القيم المحتملة التي يمكن استخراجها من الأصول التي يملكونها. ونتيجة لهذا، قد يصبحون أكثر تمسكًا بها من ذي قبل، وسيحددون أسعارًا مرتفعة للآخرين مقابل الحصول على حق الوصول إليها. واستمرارًا للتشبيه بمنجم الذهب: ستظل الأهمية القصوى في الذهب ذاته.

ولكن، هناك بُعد مهم يتعلق بالنهضة طويلة المدى لمالكي البيانات ويستحق الذكر. في بعض الحالات، سيظهر "وسطاء البيانات" الذين يستطيعون جمع البيانات من مصادرها المختلفة وتكديسها والقيام بأمور مبتكرة بها. قد يسمح مالكو البيانات لهؤلاء الوسطاء بالقيام بدورهم؛ حيث إن بعضًا من قيمة البيانات لن يأتي إلا من خلالهم.

من بين الأمثلة على هذا الأمر شركة "إنريكس"، وهي شركة تحلل حركة المرور وتقع خارج مدينة سياتل. تقوم الشركة بجمع بيانات فورية عن المواقع الجغرافية من 100 مليون مركبة في قارتي أمريكا الشمالية وأوربا. تأتيها البيانات من شركات بي إم دبليو وفورد وتويوتا من بين شركات أخرى، وكذلك من الأساطيل التجارية مثل سيارات الأجرة وشاحنات التوصيل. تحصل أيضًا على البيانات من خلال الهواتف المحولة لقائدي السيارات (حيث يكون تطبيقها المخصص للهواتف الذكية مهمًّا في هذه الحالة: يحصل المستخدمون على معلومات عن حالة المرور وفي المقابل تحصل إنريكس على إحداثياتهم). تمزج الشركة هذه المعلومات مع بيانات حالات المرور والطقس السابقة وغيرها من البيانات مثل الأحداث العامة لتوقع الحالة المرورية. تستخدم نتائج هذا المزيج الذي تنتجه الشركة لإدراجها في أنظمة الملاحة بالسيارات وتستخدمها الحكومات والأساطيل التجارية.

تعتبر شركة إنريكس وسيط البيانات المستقل الأكثر تطورًا، والتي تجمع معلوماتها من عدد كبير من شركات السيارات المتنافسة وتقوم بإنتاج منتج أكثر

قيمة من أي من المنتجات التي من المحتمل أن تنتجها هذه الشركات في يوم من الأيام. قد تمتلك كل من هذه الشركات المصنعة للسيارات عدة ملايين من النقاط البيانية التي جمعتها من سياراتها التي تعمل على الطرق. رغم أنها كانت تستطيع استخدام هذه البيانات من أجل توقع حالة المرور، فإن هذه التوقعات لم تكن لتصبح دقيقة أو كاملة تمامًا. إن جودة التوقعات تزداد بزيادة كمية البيانات المستخدمة. هذا إلى جانب أن شركات السيارات ربما لم تكن تمتلك المهارة المطلوبة: حيث إن كفاءتها تقتصر في معظمها على ثني المعادن لا التأمل في قانون بواسون للأرقام الصغيرة؛ لذا تشجعت هذه الشركات بشكل كبير لأن تستعين بأطراف أخرى من أجل القيام بالعمل. هذا إلى جانب أنه رغم أن توقعات حالة المرور مهمة للغاية بالنسبة للسائقين، فإنها لا تؤثر من قريب أو بعيد على ما إذا كان الشخص سيشتري سيارة بعينها؛ لذا فإن المتنافسين لا يمانعون أن يوحدوا جهودهم للقيام بهذا العمل.

لا شك في أن الكثير من الشركات في الكثير من المجالات قد تشاركت المعلومات فيما بينها من قبل، مثل مختبرات ضامني السندات التأمينية الشهيرة والقطاعات ذات الشبكات مثل البنوك والطاقة والاتصالات؛ حيث كان تبادل المعلومات مهمًّا لتجنب المشكلات ووضع النظم في أوقات الحاجة لها. كانت شركات أبحاث السوق تقوم بتكديس بيانات الصناعة على طوال عقود، كما فعلت الشركات ذات المهام المتخصصة مثل الشركات التي تراجع انتشار الصحف. بالنسبة لبعض الشركاء التجاريين، يعتبر هذا الأمر هو جوهر عملهم.

الفرق الذي حدث في الوقت الحالي هو أن البيانات أصبحت من المواد الخام التي تدخل إلى الأسواق، أصولاً مستقلة عما كانت تهدف لقياسه فيما مضى. على سبيل المثال، يبدو أن معلومات شركة إنريكس أكثر فائدة مما كانت تبدو عليه؛ حيث يمكن لتحليلها لحالة المرور أن تقيس صحة الاقتصاد العام لأنه يمكنها أن تقدم أفكارًا عن معدلات البطالة ومبيعات المتاجر والأنشطة الترفيهية. عندما بدأ تعافي الاقتصاد الأمريكي في الانتكاس عام 2011، تم جمع بعض من العلامات على هذا الأمر من خلال تحليل حالة المرور رغم إنكار السياسيين حدوث هذا التعافي: أصبحت أوقات الذروة أقل ازدحامًا مما يدل على المزيد من معدلات البطالة. باعت شركة إنريكس أيضًا هذه البيانات إلى شركة تعمل في مجال التمويل الاستثماري والتي تستخدم أنماط الحالة المرورية حول المتاجر الكبيرة كإشارة على مبيعاتها، والتي تستخدمها الشركة لتحديد بيع أسهمها قبل الإعلان عن توزيع الأرباح ربع السنوية؛ حيث إن المزيد من السيارات حول المتاجر يدل على المزيد من المبيعات.

هناك وسطاء آخرون ظهروا فجأة داخل سلسلة قيم البيانات الضخمة. من بين هؤلاء الوسطاء الذين ظهروا في وقت مبكر شركة هيتوايس التي اشترتها إكسبيريان فيما بعد، التي أبرمت اتفاقات مع مزودي خدمات الإنترنت للحصول على بيانات عن تدفقات ضغطات المستخدمين مقابل حصولها على دخل إضافي. تم إسناد حق استخدام البيانات لها مقابل قيمة مادية ثابتة بدلاً من الحصول على

نسبة من القيمة التي ستنتجها الشركة، وبذلك حصلت هيتوايس على أغلب القيم من خلال عملها كوسيط. مثال آخر على هذا الأمر وهو شركة كوانتكاست التي كانت تقيس معدل حركة المستخدمين عبر مواقع الإنترنت لمساعدتها على معرفة المزيد عن ديموغرافية زوارها وكذلك أنماط الاستخدام. لقد أمدت المواقع الإلكترونية بأدوات تمكنها من متابعة زيارات المستخدمين لها، وفي المقابل تتمكن كوانتكاست من الحصول على البيانات، الأمر الذي يمكنها من تحسين نظامها للإعلانات التي تستهدف قطاعًا معينًا من المستخدمين.

تمكن هؤلاء الوسطاء الجدد من الحصول على مواقع لائقة ومربحة دون تهديد النماذج العملية لمالكي البيانات الذين يحصلون على البيانات منهم. في الوقت الحالي، تعتبر الدعاية عبر الإنترنت من بين هذه المواقع اللائقة؛ حيث إنها تحتوي على أغلب البيانات وتوجد حاجة ملحة للتنقيب في هذه البيانات من أجل استهداف شرائح معينة من المجتمع بهذه الإعلانات. ولكن مع تحول المزيد مما يوجد في العالم إلى الصيغة الرقمية وإدراك المزيد من المجالات أن جوهر أعمالهم يستفيد كثيرًا من البيانات، فإن وسطاء المعلومات المستقلين هؤلاء سيظهرون في أماكن أخرى أيضًا.

قد لا يكون بعض هؤلاء الوسطاء من الشركات التجارية، بل من المؤسسات غير الهادفة للربح. على سبيل المثال، معهد تكاليف الرعاية الصحية الذي تم إنشاؤه عام 2012 على يد مجموعة من شركات التأمين الصحي الكبرى. بلغت كمية البيانات التي جمعتها هذه الشركات حوالي 5 مليارات مطالبة (مجهَّلة) تشتمل على 33 مليون شخص. مكنت مشاركة السجلات هذه الشركات من تحديد التوجهات التي لم يكونوا ليتمكنوا من رؤيتها باستخدام مجموعاتهم البيانية الفردية الصغيرة. من بين النتائج الأولية التي حصلوا عليها أن تكاليف الرعاية الطبية في الولايات المتحدة الأمريكية قد ارتفعت بمعدل ثلاث مرات أسرع من معدلات التضخم في عام 2010-20، ولكن مع وجود اختلافات واضحة على المستوى التفصيلي: ارتفعت أسعار غرف الطوارئ بمعدل 11% في حين انخفضت تكلفة المنشآت التمريضية. كان من الواضح أن شركات التأمين الصحي لم تكن لتسلم بياناتها الغالية إلى أي طرف سوى وسيط غير هادف للربح. إن دوافع المؤسسات غير الهادفة للربح لا يرتاب بها أحد، ويمكن تصميم المؤسسة مع وضع الشفافية والمحاسبة في الحسبان.

يُظهر تنوع شركات البيانات الضخمة كيفية تحول قيمة المعلومات. في حالة موقع De، تُحدد أسعار البيانات بواسطة المواقع الإلكترونية الشريكة على أساس مشاركة العائد. يحصل الموقع على عمولات مقابل شراء المستخدمين بضائع من خلاله، ولكن تحصل الشركات التي أمدت الموقع بالمعلومات أيضًا على جزء من أرباحه. يدل هذا الأمر على نضج الطريقة التي يتعامل بها هذا المجال مع البيانات: في الماضي، لم تحصل شركة ITA على أية عمولات مقابل البيانات التي أمدت فاير كاست بها، بل حصلت على مجرد مقابل ثابت للحصول على حق استخدام البيانات. ولكن في الوقت الحالي، أصبح مزودو البيانات قادرين على وضع شروط أكثر

جاذبية في اتفاقاتهم. بالنسبة للشركة الجديدة التي سيؤسسها "إتزيوني"، قد يفترض بعض الأشخاص أنه سيحاول أن يمدها بالبيانات بنفسه؛ حيث إن القيمة قد انتقلت من الخبرة إلى الفكرة وتتحرك الآن نحو البيانات ذاتها.

لقد انقلبت نماذج الأعمال رأسًا على عقب بتحول القيمة إلى هؤلاء الذين يتحكمون بالبيانات. الشركة الأوربية صانعة السيارات التي عقدت صفقة ملكية فكرية مع المورد كانت تمتلك في مكاتبها فريقًا قويًّا لتحليل البيانات ولكنها كانت بحاجة إلى التعامل مع مورد خارجي للتكنولوجيا من أجل الحصول على أفكار من هذه البيانات، وحصلت الشركة التكنولوجية على أتعاب عملها، ولكن احتفظت شركة السيارات بالأرباح كاملة. وعندما لاحت لها الفرصة، عدلت شركة التكنولوجيا من نموذج عملها لتشارك بعضًا من المخاطر والمكاسب مع عملائها؛ حيث ارتضت بالعمل مقابل أتعاب منخفضة في مقابل مشاركة بعض من الثروات التي سيكشف عنها تحليلها للبيانات. (بالنسبة لموردي قطع غيار السيارات، من المرجح أن يقولوا جميعًا في المستقبل إنهم سيرغبون في تركيب حساسات قياس في منتجاتهم، أو أن يصروا على الوصول إلى بيانات الأداء كجزء أساسي من عقود المبيعات، من أجل التطوير المستمر للمكونات التي يصنعونها).

أما بالنسبة لوسطاء البيانات، فإن حياتهم معقدة بسبب أنهم بحاجة إلى إقناع الشركات بقيمة المشاركة. على سبيل المثال، بدأت شركة إنريكس في جمع بيانات تتعدى مجرد معلومات المواقع الجغرافية. في عام 2012، أجرت تجربة لتحليل متى وأين يقوم السائقون بالضغط على أنظمة المكابح الآلية بسياراتهم، من أجل إحدى شركات السيارات التي ابتكرت نظام القياس عن بُعد الخاص بها لجمع المعلومات بشكل فوري. تدور الفكرة على أن الضغط المتواصل على المكابح على أحد الطرق يعني أن حالة هذا الطريق خطرة، وأنه على السائقين أن يفكروا في سلوك طرق بديلة. وبالتالي، وباستخدام هذه البيانات، لم تتمكن إنريكس من تقديم توصيات للسائقين بأقصر الطرق فحسب، بل بأكثرها أمانًا أيضًا.

ولكن لم تكن شركة السيارات تخطط لمشاركة هذه البيانات مع أي طرف آخر، وأصرت بدلاً من هذا على أن تدمج شركة إنريكس هذا النظام بسياراتها بشكل حصري. يبدو أن القيمة المستفادة من الإعلان عن السمات تفوق العائد من تكديس البيانات مع بيانات الآخرين من أجل تحسين الدقة العامة للنظام. يمكن القول، إن شركة إنريكس ترى أنه في وقت ما ستدرك جميع شركات السيارات فائدة المزج بين البيانات التي يملكونها جميعًا. تمتلك شركة إنريكس، كوسيط للبيانات، حافزًا قويًّا لأن تعلق آمالها على هذا التطلع: إن عملها كله يتمحور حول إمكانية وصولها إلى مصادر متعددة للبيانات.

تجرب الشركات أيضًا أنواعًا مؤسسية مختلفة في مجال البيانات الضخمة. لم تصر شركة إنريكس على استخدام نموذج عملها كما تفعل الكثير من الشركات الناشئة حديثًا ـــ حيث إن دورها كوسيط للبيانات قد تحدد منذ تصميمها. شركة مايكروسوفت التي تمتلك براءات الاختراع الرئيسية في مجال التكنولوجيا، تبينت أن إحدى الشركات المستقلة الصغيرة ـــ بدلاً من إحدى الشركات الكبرى ـــ من

شأنها أن تكون محايدة أكثر وأن تجمع بين الشركات المتنافسة في المجال الواحد وأن تحصل على أكبر كم ممكن من حقوق ملكيتها الفكرية. بالمثل، مركز ميدستار واشنطن الطبي الذي استخدم برنامج أمالجا من شركة مايكروسوفت لتحليل معدلات عودة المرضى للمستشفيات كان يعلم بدقة ما يفعله بهذه البيانات: كان برنامج أمالجا مملوكًا للمركز الطبي بالفعل كبرنامج لتشغيل غرفة الطوارئ بالمستشفى وكان يطلق عليه "آزيكسي" وبيع إلى مايكروسوفت عام 2006 التي طورته وحسنته.

في عام 2010، باعت شركة يو بي إس وحدة لتحليل البيانات طورتها داخل الشركة كانت تدعى يو بي إس للتكنولوجيا اللوجيستية، إلى شركة أسهم خاصة تسمى توما برافو. بعد أن أصبحت تعمل الآن تحت اسم رودنت للتكنولوجيا، أصبحت الوحدة أكثر حرية لتقوم بعمل تحليلات الطرق لأكثر من شركة. تقوم رودنت بجمع البيانات من عدد من العملاء لإمداد مجال الشحن بأكمله بخدمة إرشادية تستخدمها يو بي إس ومنافسوها على حدٍّ سواء. رغم أنها كانت تابعة لشركة يو بي إس، فإنها لم تحاول قط أن تقنع منافسي الشركة الأم بأن يسلموها بياناتهم، كما يشرح المدير التنفيذي للشركة "لين كينيدي". ولكن بعد أن أصبحت شركة مستقلة، أصبح منافسو يو بي إس أكثر قابلية لأن يمدوها ببياناتهم، وفي النهاية استفاد الجميع من الدقة المحسنة التي وصلنا إليها بفضل تراكم البيانات.

إن الدليل على أن البيانات ذاتها ستحمل قيمة أكبر من المهارات أو العقلية يكمن في عدد من الاستحواذات التي حدثت بين شركات البيانات الضخمة؛ فعلى سبيل المثال، في عام 2006، كافأت شركة مايكروسوفت عقلية "إتزيوني" عن البيانات الضخمة بشراء شركة فاير كاست مقابل 110 ملايين دولار، ولكن بعد عامين دفعت جوجل 700 مليون دولار مقابل الاستحواذ على مورد البيانات إلى فاير كاست، شركة ITA للبرمجيات.

وصية الخبير

في فيلم *Moneyball*، الذي يتحدث عن كيفية تحول فريق أوكلاند إلى فريق ناجح من خلال تطبيق التحليلات وأنواع القياس الجديدة على الرياضة، كان هناك مشهد رائع يصور مجموعة من كشافة اللاعبين كبار السن يجلسون حول منضدة مستديرة ويتناقشون حول اللاعبين. لم يتمكن الجمهور من تجنب الشعور بالنفور، ليس لأن المشهد يكشف عدم استخدام البيانات في المناقشة، بل لأننا جميعًا قد مررنا بمواقف كان يعتمد فيها "الاختيار" على الشعور بدلاً من العلم.

قال أحد الكشافة: "إنه يمتلك جسدًا مثاليًّا لممارسة البيسبول .... ووجهًا وسيمًا".

أكد على هذا كشاف آخر يرتدي سماعة لضعاف السمع قائلاً: "إنه يؤرجح المضرب بطريقة رائعة، وعندما يرتطم بالكرة يمكنه أن يوجهها، وتنطلق فجأة بعيدًا عن المضرب". اتفق معه كشاف آخر قائلاً: "إنها تنطلق بعيدًا جدًّا".

يقاطع رجل ثالث المحادثة قائلاً: "ولكن زوجته قبيحة".

قال كشاف اللاعبين الذي يدير الاجتماع: "وما الذي يعنيه هذا؟".

رد عليه الرجل كما لو كان يشرح حقيقة مثبتة: "الزوجة القبيحة تعني عدم الثقة بالنفس".

قال قائد الاجتماع راضيًا ومستعدًّا للانتقال لموضوع آخر: "حسنًا".

بعد بعض الدعابات الجيدة، قال أحد الكشافة الذين كانوا صامتين: "إن هذا اللاعب يمتلك أسلوبًا، أسلوبًا جيدًا. بمعنى، أنه الفتى المنشود". يضيف رجل آخر: "يمكنه أن يمرر الكرة بشكل جيد، ويتمتع بالوسامة، وهو جاهز للعب دوره في الفريق ولكنه يحتاج إلى اللعب لوقت أطول".

يتدخل الرجل الذي قاطعهم في المرة الأولى قائلاً: "إنني أقول إن شكل زوجته لا يهم كثيرًا على ما أعتقد".

يصور المشهد عيوب الحكم البشري. إن جميع ما تمت مناقشته في هذا الجدل لا يعتمد على أي شيء ثابت؛ حيث إن اتخاذ قرارات بتوقيع عقود للاعبين بملايين الدولارات يتم اتخاذها بناءً على الحدس، دون اللجوء إلى المعايير الموضوعية. نعم، إنه مجرد فيلم، ولكن الحياة الواقعية لا تختلف كثيرًا؛ حيث يتم توظيف المنطق الخاوي نفسه من مكاتب مجالس إدارة الشركات في مانهاتن إلى المكتب البيضاوي بالبيت الأبيض إلى المقاهي ومناضد المطابخ بالمنازل وكل مكان.

فيلم *Moneyball*، المقتبس من أحد كتب "مايكل لويس"، يروي القصة الحقيقية للمدرب "بيلي بين"، مدرب الفريق الأول بنادي أوكلاند الذي ألقى بكتاب قواعد تقييم اللاعبين القديم من أجل طريقة تعتمد على الرياضيات والتي تتناول رياضة البيسبول من خلال مجموعة جديدة من القياسات، والتي تخلت عن بعض الإحصائيات القديمة مثل "معدل الضربات" واستعانت بطرق تفكير قد تبدو غريبة عن اللعبة مثل "نسبة الوقوف على القواعد". كشفت الطريقة المعتمدة على البيانات عن بعد جديد من الرياضة كان حاضرًا طوال الوقت ولكنه مختفٍ بين الكثير من الأمور المزدحمة. لا تهم الكيفية التي وصل بها اللاعب إلى القاعدة سواء من خلال التقاط الكرة المتدحرجة على الأرض أو من خلال سير الهوينى، طالما أنه قد وصل إليها. عندما اكتشفت البيانات أن سرقة القواعد أمر غير فعال، قامت بالتخلي عن عناصر اللعبة الأكثر تشويقًا بعد أن أثبتت أنها الأقل "فائدة".

في خلال واحدة من أكثر المجادلات أهمية، خزن "بين" في مكتب استقبال الفريق الأسلوب الذي عُرف باسم إحصائيات السيف، المصطلح الذي صاغه الكاتب الرياضي "بيل جايمس" في إشارة إلى مجتمع أبحاث رياضة البيسبول الأمريكي، والذي كان يعتبر حتى ذلك الوقت مقصورًا على مجموعة من القلة المهووسين بتلك الرياضة. كان "بين" يتحدى عقيدة الخندق، كما تحدت وجهات نظر جاليليو عن مركزية الشمس سلطة الكنيسة الكاثوليكية. وفي نهاية الأمر تمكن من قيادة الفريق الذي عانى طويلاً ليحرز المركز الأول في دوري الغرب الأمريكي في موسم 2002، مع إحراز 20 فوزًا متتاليًا دون هزيمة. منذ ذلك الوقت، حل الإحصائيون محل كشافة اللاعبين حيث أصبحوا عباقرة الرياضة. وتوجهت الكثير من الفرق الرياضية للاستعانة بإحصائيات السيف هي الأخرى.

من المنطلق ذاته، سيكون التأثير الأكبر للبيانات الضخمة هو أن القرارات المبنية

على البيانات تميل إلى الزيادة من الحكم البشري أو الانتقاص منه. في كتابه تحت عنوان Super Crunchers، يقول عالم الاقتصاد بجامعة يال وأستاذ القانون "إيان أيرز" إن التحليل الإحصائي يجبر الناس على وضع حدسهم في الاعتبار، ومع البيانات الضخمة تزداد أهمية هذا الأمر أكثر فأكثر. سوف يفقد خبراء الدراسات الموضوعية والمتخصصون الواقعيون بعضًا من بريقهم مقارنة بالإحصائيين ومحللي البيانات، الذين تحرروا من قيود الطرق التقليدية القديمة للقيام بالأمور والذين يسمحون للبيانات بأن تتحدث. سيعتمد المجال الوظيفي الجديد هذا على العلاقات التبادلية دون اللجوء إلى الحكم المسبق أو الانحياز إلى أحد الأطراف، تمامًا كما فعل "موراي" عندما لم يتقبل جميع القيم الاسمية التي كان يتحدث عنها قادة السفن القدامي عندما كانوا يمرحون فيما بينهم، حيث إنه لم يثق إلا بالبيانات المتراكمة التي تكشف حقائق عملية.

لقد شهدنا أفول نجم تأثير خبراء الدراسات الموضوعية في الكثير من المجالات؛ حيث إن المحتوى المنشور في وسائل الإعلام مثل هافينجتون بوست وجاوكر وفوربس أصبح يتحدد تدريجيًّا بالبيانات، وليس بحكم المحررين من البشر وحدهم؛ حيث إن البيانات بمقدورها أن تكشف عما يرغب الناس في قراءته أفضل من حدس الصحفيين المتمرسين. تستخدم شركة التعليم عبر الإنترنت، "كورسيرا"، البيانات لتعرف أي قسم من أقسام أحد المحاضرات المصورة يعيد الطلبة مشاهدته مرة بعد أخرى لتعرف أيًّا من مواد مناهجها لا يتم شرحها بشكل كافٍ، ومن ثم تعيد تلك المعلومات إلى المعلمين كي يحسنوا من أسلوب شرحهم. كما قلنا من قبل، طرد "جيف بيزوس" الكثير من مراجعي الكتب من البشر من شركة أمازون بعدما تبين له أن التوصيات القائمة على معادلات الخوارزميات تزيد من نسبة المبيعات.

يعني هذا الأمر أن المهارات المطلوبة للنجاح في سوق العمل قد تغيرت، فقد تغيرت الأمور التي من المتوقع أن يفيد بها الموظفون مؤسساتهم. الدكتورة "ماكجريجور" والتي ترعى الأطفال حديثي الولادة المبتسرين في أونتاريو، والتي لا تحتاج إلى أن تكون أكثر أطباء المستشفى معرفة، أو أفضل أطباء العالم في رعاية الأطفال حديثي الولادة، حتى تحقق أفضل النتائج الصحية لمرضاها. في حقيقة الأمر، إنها ليست طبيبة معالجة على الإطلاق؛ حيث إنها تحمل درجة الدكتوراه في علوم الحاسب. ولكنها تفيد نفسها بكم كبير من البيانات مكدسة منذ أكثر من عقد من الزمان من سجلات المرضى، والتي يحللها الحاسب وتستخدم النتائج في تقديم توصيات بطرق العلاج المناسبة.

رأينا أن رواد مجال البيانات الضخمة عادة ما يأتون من مجالات أخرى غير المجال الذي يضعون بصمتهم به؛ حيث يكونون من المتخصصين في مجالات تحليل البيانات أو الذكاء الاصطناعي أو الرياضيات أو الإحصاء، ومن ثم يطبقون هذه المهارات على مجالات بعينها. عادة ما يكون الفائزون بمسابقات "كاجل"، منصة الإنترنت لمشروعات البيانات الضخمة، من الوافدين الجدد على هذا المجال والذين يتمكنون من تحقيق نتائج جيدة، كما يقول "أنتوني جولدبلوم" المدير التنفيذي لشركة

"كاجل". قام أحد علماء الطبيعيات البريطانيين بتطوير معادلة اقتربت من الفوز بالجائزة لتوقع الادعاءات التأمينية وتحديد السيارات المستعملة المعيبة. وقام أحد الخبراء بشئون التأمين من سنغافورة بعمل مسابقة لتوقع ردود الفعل البيولوجية على المركبات الكيميائية المختلفة. في الوقت ذاته، في مجموعة جوجل للترجمة الآلية، احتفل المهندسون بتمكنهم من ترجمة لغات لا يتحدثها أحد في مكتب الشركة. وبالمثل، كان الإحصائيون في وحدة الترجمة الآلية بشركة مايكروسوفت يتداولون دعابة قديمة: إن جودة الترجمة تتزايد عندما يغادر أحد علماء اللغة فريق العمل.

كن متأكدًا من أن خبراء الدراسات الموضوعية لم ينقرضوا تمامًا، ولكن سينحسر تميزهم. من الآن فصاعدًا، سيكون لزامًا عليهم أن يتشاركوا منصة التتويج مع عباقرة البيانات الضخمة، كما يجب على العلاقات السببية المتوجة أن تتشارك الأضواء مع العلاقات التبادلية المتواضعة. يغير هذا الأمر من الطريقة التي نقدر بها المعرفة؛ حيث إننا نميل إلى الاعتقاد بأن الأشخاص الذين يتخصصون في مجال معين يستحقون تقديرًا أكبر من الممارسين العامِّين ــ أن الثروة تفضل التعمق. ولكن ما زالت التخصصات مشابهة لفكرة الدقة التي تحدثنا عنها من قبل: تلائم عالم البيانات الصغيرة الذي لا يمكن للمرء أبدًا أن يحصل فيه على كم كافٍ من المعلومات، أو المعلومات المناسبة، وبالتالي يجب أن يعتمد على الحدس والخبرة ليستمر في العمل. في مثل هذا العالم، تلعب الخبرة دورًا مهمًّا؛ حيث إنها التراكم الطويل للمعرفة الكامنة في العقول ــ المعرفة التي لا يمكن للمرء أن يحصل عليها بسهولة أو أن يتعلمها نظريًّا من الكتب، أو ربما يكون مدركًا وجودها ــ وهو الأمر الذي يساعد المرء على اتخاذ القرارات الصائبة.

عندما تمتلك كمًّا كبيرًا من البيانات، يمكنك أن تستغلها جيدًا، وأن تحصل منها على نتائج عظيمة؛ لهذا فإن الأشخاص القادرين على تحليل البيانات الضخمة يمكنهم أن يتخطوا الخرافات وطرق التفكير التقليدية ليس لأنهم أكثر ذكاءً، بل لأنهم يملكون البيانات التي تساعدهم على ذلك. (ولأنهم من الوافدين الجدد على المجال، فسيكونون محايدين فيما يتعلق بالخصومات التي تحدث داخله والتي قد تؤثر سلبًا على وجهة نظر الخبراء فيما يتعلق بأية جهة سيناصرون في هذه الخصومة). ويقترح هذا الأمر ما يحتاج إليه الموظف ليكون عضوًا مؤثرًا في تغير أسلوب عمل الشركة. إن ما تحتاج إلى معرفته قد تغير، ومن تحتاج إلى معرفته قد تغير، وكذلك تغير ما تحتاج إلى دراسته من أجل الاستعداد لحياتك المهنية.

الرياضيات والإحصاء، إلى جانب القليل من علوم برمجة الحواسب والشبكات، ستكون من أساسيات سوق العمل كما كانت الحسابات منذ قرن مضى ومعرفة القراءة والكتابة في العصور السابقة لذلك. في الماضي، إذا ما كنت ترغب في أن تصبح عالمًا لامعًا في الأحياء، كان لزامًا عليك أن تتعلم الكثير من علماء الأحياء الآخرين. ولم يتغير هذا الأسلوب بالكامل. ولكن تعمل البيانات الضخمة على زيادة سعة الموضوعات إلى جانب زيادة عمق التخصص فيها. يمكن حل المشكلات الإحيائية المستعصية بالتعاون مع عالم بالفلك الطبيعي أو مصمم للبيانات البصرية.

يعتبر مجال ألعاب الفيديو من المجالات التي قام فيها خبراء البيانات الضخمة بشق طريقهم لينافسوا رواد هذا التخصص، من خلال تغيير طريقة عمل المجال بأكمله. إن قطاع ألعاب الفيديو يعتبر من الأعمال الكبرى؛ حيث توجد به أرباح تتخطى الأرباح السنوية لشباك تذاكر أفلام هوليوود في جميع أنحاء العالم. في الماضي، كانت الشركات تصمم لعبة ما، وتصدرها في الأسواق، وتنتظر آملة أن تنجح نجاحًا باهرًا. وعلى أساس بيانات المبيعات، تقرر الشركة ما إذا كانت ستحول اللعبة إلى سلسلة أم ستبدأ مشروعًا جديدًا. كانت القرارات المتخذة فيما يتعلق بسرعة اللعب وعناصر اللعبة من شخصيات ومغامرات وأغراض وأحداث تعتمد على ابتكار المصممين وحدهم، والذين كانوا يهتمون بعملهم بدرجة اهتمام مايكل أنجلو بعمله في أثناء ما كان يرسم سقف سيستين بالفاتيكان. كان الأمر بالنسبة لهم فنًّا وليس علمًا، عالمًا يحركه الحدس والمواهب، يشبه إلى حدٍّ كبير حالة كشافة اللاعبين في فيلم *Moneyball*.

لقد ولت هذه الأيام جميعها. لقد أصبحت ألعاب شركة زينجا مثل فارم فيل وفرونتيير فيل وفيش فيل وغيرها موجودة على شبكة الإنترنت وأصبحت كذلك تفاعلية. للوهلة الأولى، قد ترى أن ممارسة الألعاب عبر الإنترنت تسمح لشركة زينجا باستعراض بيانات الاستخدام وتعديل الألعاب على أساس كيفية ممارسة المستخدمين لها؛ لذا فإن كان المستخدمون يواجهون صعوبة في الانتقال من مستوى لآخر، أو يميلون إلى ترك اللعبة في وقت ما لأن اللعبة فقدت متعتها، يمكن للشركة أن تدرك هذا الأمر من خلال البيانات وتصلح الوضع على الفور. ولكن الأمر غير الواضح تمامًا هو أن الشركة يمكنها أن تصمم ألعابًا تناسب صفات مجموعات معينة من المستخدمين. لا توجد نسخة واحدة من لعبة فارم فيل، بل توجد مئات النسخ منها.

يدرس محللو البيانات الضخمة بشركة زينجا ما إذا كانت مبيعات البضائع الافتراضية تتأثر بلونها أم برؤية المستخدم لأصدقائه وهم يستخدمونها. على سبيل المثال، بعدما كشفت البيانات عن أن لاعبي فيش فيل قد قاموا بشراء سمكة نصف شفافة بمعدل يعادل 6 مرات شراء المخلوقات الافتراضية الأخرى، قامت زينجا بتصميم المزيد من المخلوقات الافتراضية نصف الشفافة وحصلت على أرباح ممتازة. في لعبة حروب المافيا، كشفت البيانات أن اللاعبين قاموا بشراء المزيد من الأسلحة ذات الحواف الذهبية واشتروا أيضًا النمور الأليفة ذات اللون الأبيض.

هذه الأمور من النوعية التي لم يكن بمقدور المصممين القابعين في الأستوديوهات إدراكها، ولكن قالت البيانات كلمتها. يصف "كين رودين"، رئيس قسم الإحصاء بشركة زينجا، قبل أن ينتقل منها ليكون رئيس قسم الإحصاء بشركة فيس بوك، شركة زينجا قائلاً: "إننا شركة إحصائيات متخفية في صورة شركة ألعاب. إن كل شيء يدار هنا بالأرقام". إن توظيف البيانات لا يضمن نجاح الشركات ولكنه يكشف لها احتمالات الحصول على أرباح أكثر.

إن الانتقال إلى اتخاذ القرارات باستخدام البيانات أمر مهم. إن أغلب الأشخاص يبنون قراراتهم على أساس المزج بين الحقائق والتفكير، إلى جانب قدر كبير من

التخمين. كما يقول الشاعر "دبليو إتش أودين" في كلماته الخالدة "يشبه تزاحم الرؤى الموضوعية المشاعر في أشعة الشمس الذهبية". يطلق "توماس دافنبورت"، أستاذ إدارة الأعمال بكلية بوسطن بولاية ماساتشوسيتس ومؤلف العديد من الكتب عن الإحصاء، على هذا الأمر "الغريزة الذهبية". يستمد المديرون التنفيذيون ثقتهم من حدسهم الغريزي؛ لذا فإنهم يواصلون العمل من خلاله. ولكن بدأت جميع هذه الأمور في التغير منذ أن بدأ اتخاذ القرارات الإدارية أو على الأقل يتم التأكيد عليها بواسطة نماذج التوقعات وتحليل البيانات الضخمة.

على سبيل المثال، يستخدم موقع The-Numbers.com الكثير من البيانات والإحصاءات ليخبر منتجي هوليوود المستقلين بالدخل الذي قد يتحصل عليه أحد الأفلام قبل البدء بتصويره. تقوم قاعدة بيانات الشركة بتحليل ما يزيد على 30 مليون سجل تغطي جميع الأفلام التجارية في الولايات المتحدة الأمريكية لعقود ماضية؛ حيث تشمل ميزانيات جميع الأفلام ونوعها الأدبي والممثلين وأطقم العمل والجوائز التي حصلت عليها وكذلك الدخل الذي تحصلت عليه (من شبابيك التذاكر بالولايات المتحدة الأمريكية ومن جميع أنحاء العالم، وشراء الحقوق في قارات أخرى ومبيعات نوادي الفيديو وإيجارات الأفلام وغيرها)، وغيرها الكثير. تحتوي قاعدة البيانات أيضًا على شبكة من الصلات البشرية، مثل "هذا الكاتب السينمائي عمل مع هذا المخرج، هذا المخرج عمل مع هذا الممثل"، كما يقول مؤسس الشركة ورئيسها "بروس ناش".

يمكن لشركة The-Numbers.com أن تكتشف العلاقات التبادلية المعقدة التي يمكنها توقع الدخل الذي يمكن أن تحققه المشروعات السينمائية. ويأخذ المنتجون هذه المعلومات إلى شركات الإنتاج أو المستثمرين ليحصلوا على التمويل. يمكن للشركة حتى أن تقوم بتعديل المتغيرات من أجل إخبار عملائها بالطرق التي تساعدهم على زيادة المساعدات التي يحصلون عليها (أو التقليل من الخسائر). في إحدى الحالات، أظهرت تحليلاتها أن فرص نجاح المشروع السينمائي ستزداد إذا ما قام بالدور الرئيسي للذكور في الفيلم ممثل من نجوم الصف الأول: خاصة إذا ما كان من الفائزين بالأوسكار ويتقاضى أجرًا حوالي 5 ملايين دولار. في حالة أخرى، أخبر "ناش" شركة إيماكس للإنتاج السينمائي أن فيلمها التسجيلي عن الإبحار الشراعي قد يكون مربحًا إذا ما قللت ميزانيته من 12 مليون دولار إلى 8 ملايين دولار. يقول ناش: "لقد جعل هذا الأمر المنتج سعيدًا ـــ أما المخرج فلم يكن على هذا القدر من السعادة".

من إنتاج الأفلام إلى تحديد مواقع تمركز لاعبي البيسبول، بدأ التحول في اتخاذ القرارات المؤسسية يظهر على النتائج. درس "إيريك برينجولفسون"، أستاذ إدارة الأعمال بمدرسة إم آي تي سلون لإدارة الأعمال، بالتعاون مع مجموعة من زملائه أداء الشركات المتميزة في اتخاذ القرارات المؤسسية بناءً على تحليل البيانات ومقارنته بأداء الشركات الأخرى التي لا تفعل ذلك، ووجدوا أن مستوى إنتاجية هذه الشركات يزيد بنسبة 6 % عن الشركات التي لا تستعين بالبيانات في التأكيد على القرارات المؤسسية. يمنح هذا الأمر الشركات التي تستخدم البيانات ميزة إضافية

ـــ والتي تشبه ميزة عقلية ومهارات البيانات الضخمة؛ حيث إنها لن تستمر طويلاً باستمرار شركات أخرى في الاستعانة بطرق البيانات الضخمة لتطوير أعمالها.

مسألة الفائدة

لقد أصبحت البيانات الضخمة مصدرًا للمنافسة بين الكثير من الشركات؛ حيث إن بنية مجالات كاملة سيعاد تشكيلها من جديد. ولكن لن تحصل الشركات على الفوائد من هذا الأمر بشكل متساو. وسيكون الفائزون من بين الشركات الكبرى والصغرى التي ستحصر أغلبية الشركات الأخرى في المنتصف فيما بينها.

ستواصل الشركات الكبرى مثل جوجل وأمازون تحليقها بعيدًا عن الآخرين. وعلى العكس من الوضع في العصر الصناعي، لن يقتصر تفوقهم التنافسي على المعيار المادي. إن البنى التحتية الضخمة لمراكز البيانات التي تديرها هذه الشركات مهمة للغاية ولكنها ليست العنصر الأساسي في جودتها. مع وجود وفرة كبيرة في مساحات التخزين الرقمية وآليات المعالجة المتوافرة للإيجار بأسعار منخفضة والتي يمكن الحصول عليها في غضون دقائق معدودة، يمكن للشركات أن تعدل من عدد أجهزة الحاسب والمساحات التخزينية التي تملكها لتتوافق مع متطلباتها الحقيقية. من خلال تحويل ما كان يُعرف بأنه تكلفة ثابتة إلى متغيرة، سيقلل هذا التغيير من التفوق القياسي القائم على البنى التكنولوجية التحتية التي طالما تمتعت بها الشركات الكبرى.

إن القياس لا يزال مهمًّا، ولكنه تغير. إن ما يهم الآن هو قياس البيانات. يعني هذا الأمر امتلاك كميات هائلة من البيانات والقدرة على الحصول على المزيد منها بيسر وسهولة. ستزدهر هذه الشركات التي تملك كميات كبيرة من البيانات من خلال جمع المزيد من المواد الخام في نماذج أعمالها وتخزينها، والتي يمكنها إعادة استخدامها من أجل الحصول على قيم مادية إضافية.

إن التحدي الذي يواجه الفائزين في عالم البيانات الصغيرة وأبطال العمل خارج الإنترنت ـــ شركات مثل وولمارت وبروكتور وجامبل وجي إي ونستله وبوينج ـــ هو تقدير قوة البيانات الضخمة وجمعها واستخدامها بطريقة أكثر إستراتيجية. قامت الشركة المصنعة لمحركات الطائرات "رولز رويس" بتغيير نموذج عملها تمامًا خلال الأعوام العشرة الماضية من خلال تحليل البيانات التي جمعتها من منتجاتها، وليس مجرد صناعتها. من مركز عملياتها في بريطانيا، تتابع الشركة باستمرار أداء ما يزيد على 3700 محرك طائرة نفاثة على مستوى العالم لتحديد الأعطال قبل حدوثها. استخدمت الشركة البيانات لمساعدتها على تحويل نفسها من شركة مصنعة إلى شركة مزدوجة المهام: تبيع الشركة المحركات ولكنها تعرض على عملائها أن تتابع أداءها، وتحمل العميل التكاليف طبقًا لفترة استخدام هذه الخدمة (وتصلحها أو تستبدلها في حالة وجود خطب ما). تمثل الخدمات في الوقت الحالي نسبة 70 % من دخل قطاعات محركات الطائرات المدنية كل عام.

الشركات الجديدة، مثلها مثل الشركات القديمة في مجالات العمل الجديدة، تحاول الحصول على موقع متميز في مجال العمل من أجل الحصول على كميات

كبيرة من البيانات. غزت شركة أبل عالم الهواتف المحمولة لهذا الغرض. قبل ظهور الآي فون، كانت شركات شبكات الهواتف المحمولة تجمع كميات ضخمة من بيانات مشتركيها ولكنها فشلت في الاستفادة منها. أما أبل فعلى النقيض؛ حيث إنها وضعت في عقودها مع شركات شبكات الهواتف المحمولة شروطًا تقضي بحصولها على المعلومات الأكثر فائدة. من خلال الحصول على البيانات من عدد من شركات الشبكات اللاسلكية من جميع أنحاء العالم، حصلت شركة أبل على صورة أكثر شمولاً عن استخدام الهواتف المحمولة أكثر من أية شركة شبكات لا سلكية وحدها.

تقدم البيانات الضخمة أيضًا فرصًا رائعة للشركات التي على الجانب الآخر من مقياس الحجم؛ حيث يمكن للشركات الصغيرة الماهرة والذكية أن تتمتع بوجود "مقياس دون حجم كبير"، طبقًا للعبارة الشهيرة التي أطلقها الأستاذ "برينجولفسن"، والتي تعني أن تلك الشركات الصغيرة يمكنها أن تمتلك حضورًا فعليًّا كبيرًا دون الكثير من الموارد المادية، ويمكنها نشر الكثير من الابتكارات على مدى واسع بتكاليف محدودة للغاية. الأمر المهم هو أن الكثير من الخدمات التي تعتمد على البيانات الضخمة تقوم بشكل أساسي على الأفكار المبتكرة، فإنها لا تتطلب الكثير من الاستثمارات في البداية. يمكن للشركات الصغيرة أن تحصل على حقوق استغلال البيانات بدلاً من امتلاكها، ويمكنها إجراء تحليلاتها من خلال منصات حوسبة سحابية منخفضة التكلفة، وأن تدفع مقابل حق استغلال البيانات في صورة نسبة من الدخل الذي ستحصل عليه منها.

هناك فرصة جيدة ألا تقتصر مميزات طرفي المقياس على مستخدمي البيانات فحسب، بل قد تمتد لتشمل مالكيها أيضًا. إن مالكي البيانات الكبار لديهم محفزات كبيرة لإضافة المزيد من البيانات إلى البيانات التي تخزنها؛ حيث إن القيام بهذا يمدها بمكاسب كبيرة في مقابل تكاليف هامشية. فأولاً، تمتلك تلك الشركات البنية التحتية للتخزين والمعالجة بالفعل. ثانيًا، يمكن الحصول على قيم خاصة من المزج بين المجموعات البيانية. ثالثًا، وجود متجر واحد للحصول على الكثير من البيانات سيسهل الأمر على مستخدمي البيانات.

ولكن قد يظهر نوع جديد من مالكي البيانات على الساحة: الأفراد. مع تزايد وضوح قيمة البيانات، قد يرغب الناس في تجربة حظوظهم كمالكين للمعلومات التي تخصهم ــ على سبيل المثال، مفضلاتهم التسوقية، عادات استعراض وسائل الإعلام، وربما البيانات الصحية أيضًا.

يمكن لامتلاك الأفراد للبيانات أن يعزز من المستهلكين الفرديين بطرق لم تكن لتحدث من قبل. قد يرغب الأشخاص في أن يتخذوا القرار بأنفسهم فيما يتعلق بمن عليهم أن يعطوا له حق استغلال بياناتهم، وتحديد المقابل المادي لهذا. من المؤكد أن تجد أشخاصًا لا يهمهم إعطاء بياناتهم لمن يدفع أكثر؛ حيث سيسعد الكثيرون برؤية بياناتهم يعاد استخدامها بالمجان مقابل الحصول على خدمات أفضل، مثل توصيات كتب أكثر دقة من أمازون وتجارب مستخدمين أفضل في موقع بينترست، وخدمة اللوحات الإعلانية ومشاركة محتويات الإنترنت. ولكن بالنسبة لعدد من المستهلكين الذين يمتلكون معرفة جيدة بالأمور الرقمية، قد تبدو فكرة تسويق

وبيع معلوماتهم الشخصية طبيعية مثلها مثل كتابة المدونات أو التغريدات أو تحرير إحدى المشاركات في موسوعة ويكيبيديا.

حتى ينجح هذا الأمر فهناك المزيد نحتاج إليه أكثر من مجرد التحول في مفضلات وممارسات المستهلكين. ربما كان في الوقت الحالي من الصعب والمكلف أن يسند الناس حق استغلال بياناتهم الشخصية للشركات ومن الصعب على الشركات أن تتواصل مع كل شخص للحصول على بياناته. ولكن الأمر الأكثر ترجيحًا، أننا سنشهد ميلاد شركات جديدة تجمع كمًّا كبيرًا من البيانات من الكثير من المستهلكين مما يمد بطريقة سهلة لإسناد حق استغلال البيانات الشخصية لآخرين، وجعل التعاملات مع المستهلكين آلية تمامًا. إذا كانت تكاليف القيام بهذا منخفضة بدرجة كافية، وإذا كان هناك الكثير من الأشخاص يثقون بهم، فمن المحتمل جدًّا أن يتم إنشاء سوق لتبادل البيانات الشخصية. هناك شركات على غرار مايدكس في بريطانيا ومجموعات مثل آي دي ثري، التي شاركت في إنشائها "ساندي ينتلاند" خبيرة تحليل البيانات الشخصية في معهد ماساتشوسيتس للتكنولوجيا، بدأت بالفعل في محاولة لتحويل هذه الرؤية إلى حقيقة.

حتى ينهض هؤلاء الوسطاء ويبدءوا في العمل ويبدأ مستخدمو البيانات في استخدام البيانات الشخصية التي سينتجونها، فإن الأشخاص الذين يرغبون في أن يكونوا مالكي بياناتهم الشخصية لا تتوافر لهم الكثير من الخيارات. خلال هذه الفترة المؤقتة، ولكي يحتفظوا بخياراتهم المتاحة حتى تقوى شكيمة الوسطاء والبنى التحتية الخاصة ببياناتهم، على الأشخاص أن يفشوا بأقل القليل من بياناتهم الشخصية.

بالنسبة للشركات المتوسطة، عادة ما تكون البيانات الضخمة أقل فائدة؛ حيث إن هناك مزايا متعلقة بالحجم بالنسبة للشركات الكبرى ومزايا تتعلق بالتكلفة والابتكار للشركات صغيرة الحجم، كما يقول "فيليب إيفانز" من المجموعة الاستشارية ببوسطن، عالم ومفكر في أمور التكنولوجيا وإدارة الأعمال. في القطاعات التقليدية، توجد الشركات متوسطة الحجم بسبب دمجها للحد الأدنى من المقياس لكي تجني فوائد المقياس مع وجود كم معين من المرونة التي تفتقدها الشركات الكبرى. ولكن في عالم البيانات الضخمة لا يوجد حد أدنى للمقياس الذي على الشركة أن تصل إليه لتدفع مقابلاً لاستثماراتها في إنتاج البنية التحتية. إن مستخدمي البيانات الضخمة الذين يرغبون في الاحتفاظ بمرونتهم ونجاحهم سيجدون أنهم لم يعودوا بحاجة إلى الوصول إلى حد معين لحجم الشركة، بل يمكن للشركة أن تظل صغيرة وتزدهر في الوقت ذاته (أو يتم الاستحواذ عليها من قبل أحد عمالقة البيانات الضخمة).

تضغط البيانات الضخمة على الشركات التي تقع في المواقع المتوسطة من هذا المجال؛ مما يدفع الشركات لأن تصبح من الشركات الكبرى، أو أن تكون من الشركات الصغيرة السريعة، أو أن تخرج تمامًا من هذا المجال. سوف تتحول الكثير من القطاعات التقليدية لتصبح من الشركات الكبرى في عالم البيانات الضخمة، من الخدمات المالية إلى الأدوية إلى الصناعة. لن تقوم البيانات الضخمة بإخراج

جميع الشركات متوسطة الحجم من المنافسة في جميع المجالات، ولكنها من المؤكد أن تضغط على الشركات في المجالات التي تكون عرضة للتأثر بقوة البيانات الضخمة.

تميل البيانات الضخمة إلى الإخلال بالتوازن بين المميزات التنافسية بين الدول أيضًا. في الوقت الذي دخلت فيه الدول النامية المنافسة في مجال الصناعة وأصبح الابتكار متوافرًا للجميع، احتفظت الدول الصناعية بميزتها التنافسية من خلال الاحتفاظ بالبيانات وإدراك كيفية استخدامها. الخبر السيئ أن هذه الميزة لن تستمر طويلاً، فكما حدث مع مجالات الحوسبة والإنترنت، سيتناقص تقدم الغرب المبكر في مجال البيانات الضخمة مع إدراك أجزاء أخرى من العالم كيفية استخدام هذه التكنولوجيا. الخبر الجيد للشركات المسيطرة على المجال في الوقت الحالي والتي تعمل في الدول المتقدمة، أن البيانات الضخمة من شأنها أن تزيد من مواطن القوة والضعف للمؤسسات على حد سواء؛ لذا إن برعت إحدى الشركات في استخدام البيانات الضخمة، فستلوح لها الفرصة ليس فقط لأن تزيد من كفاءتها، بل لأن تزيد من تقدمها على منافسيها أيضًا.

لقد بدأ السباق. وكما تحتاج معادلات البحث من شركة جوجل إلى بواقي بيانات المستخدمين لتعمل بشكل جيد، وكما رأى مورد قطع غيار السيارات الألماني أهمية البيانات لتحسين منتجاته، فإن جميع الشركات يمكنها أن تستفيد من خلال استغلال البيانات بطرق ماهرة.

رغم الفوائد الجمة للبيانات الضخمة فإن هناك أمورًا يجب أن نقلق منها. منذ أن زادت البيانات الضخمة من دقة التوقعات عن العالم وعن موقعنا فيه، فربما لا نكون على استعداد لمواجهة تأثيرها على خصوصيتنا أو على إحساسنا بالحرية. إن إدراكنا وعاداتنا بنيت من أجل عالم يعتمد على ندرة المعلومات لا على وفرتها. سنقوم باستعراض الجانب المظلم من البيانات الضخمة في الفصل التالي.

8
المخاطر

على مدى أربعين عامًا، وحتى انهيار جدار برلين عام 1989، كانت وكالة الأمن القومي لدولة ألمانيا الشرقية والتي كان يطلق عليها "ستازي" تتجسس على ملايين البشر؛ حيث وظفت حوالي مائة ألف من الموظفين الذين يعملون بدوام كامل، وتمكنت من مراقبة الناس من السيارات والشوارع. كانت تقوم بفتح المراسلات بين الناس والتجسس على حساباتهم البنكية، وقامت بتركيب أجهزة التنصت في الشقق وتجسست على خطوط الهاتف. وأغرت الأزواج والآباء والأبناء بالتجسس على بعضهم الآخر ضاربين بعرض الحائط مظاهر الثقة الرئيسية التي يتمتع بها البشر تجاه بعضهم. وكانت الملفات الناتجة عن هذا التجسس ــ وتشمل 3 مليون بطاقة فهرسة على الأقل وحوالي 70 ميلًا من المستندات ــ تسجل وتفصل أكثر المظاهر الحميمية في حياة المواطنين العاديين. كانت ألمانيا الشرقية واحدة من أكثر الدول شاملة المراقبة في التاريخ.

بعد عشرين عامًا من سقوط ألمانيا الشرقية، تم جمع المزيد من البيانات عن كل إنسان أكثر مما مضى. إننا نقع قيد المراقبة الدائمة: عندما نستخدم بطاقاتنا الائتمانية للتعامل المادي، أو هواتفنا المحمولة للتواصل مع الآخرين أو أرقام التأمين الاجتماعي لتحديد هوياتنا. في عام 2007 تشدقت وسائل الإعلام البريطانية بالسخرية عن وجود ما يزيد على 30 كاميرا مراقبة في إطار 200 ياردة من الشقة بمدينة لندن التي كتب فيها "جورج أورويل" روايته تحت عنوان 1984. قبل ظهور الإنترنت، قامت الشركات المتخصصة مثل إكويفاكس وإكسبريان وأكسيوم بجمع وجدولة وتوفير مدخل إلى المعلومات الشخصية لمئات الملايين من البشر من جميع أنحاء العالم. جعلت شبكة الإنترنت من تتبع هذه المعلومات أسهل وأرخص وأكثر فائدة، وأدت إلى حقيقة أن الوكالات الحكومية الخفية التي تتكون أسماؤها من ثلاثة أحرف ليست وحدها من يتجسس علينا. إن شركة أمازون تتابع مفضلاتنا التسوقية وتتابع جوجل عادات بحثنا على محرك البحث الخاص بها، في حين تعلم تويتر ما يدور بأذهاننا. يبدو أن شركة فيس بوك تمتلك جميع هذه المعلومات هي الأخرى، إلى جانب علاقاتنا الاجتماعية. وتعلم شركات شبكات الهواتف المحمولة، ليس فقط من نتحدث معه، بل من يوجد بالقرب منا أيضًا.

مع الوعد الذي تقدمه البيانات الضخمة بحصول من يحللها على معارف قيمة، يبدو أن جميع المؤشرات تشير إلى ارتفاع موجة جمع وتخزين وإعادة استخدام الآخرين لبياناتنا الشخصية. سيزداد حجم ومقياس جمع البيانات من خلال قفزات يسببها انخفاض تكاليف تخزين البيانات وزيادة قوة أدوات تحليلها. إذا ما كان عصر الإنترنت يهدد الخصوصية، فهل تعرضها البيانات الضخمة لمزيد من الخطر؟ هل هذا هو الجانب المظلم من البيانات الضخمة؟

نعم، إنه كذلك، ولكنه ليس الجانب المظلم الوحيد. هناك أيضًا جانب مهم من البيانات الضخمة هو أن تغيير المقياس يؤدي إلى تغيير الحالة. كما سنشرح لاحقًا، لا يجعل هذا التحول من حماية الخصوصية أمرًا صعبًا فحسب، ولكنه يشكل تهديدًا جديدًا تمامًا: الأخطاء المعتمدة على التوجهات، أي احتمالية استخدام توقعات البيانات الضخمة في الحكم على الأشخاص ومعاقبتهم قبل حتى أن يقدموا على أي فعل. إن القيام بهذا الأمر يلغي أفكار العدل والإنصاف والإرادة الحرة.

إلى جانب الخصوصية والتوجهات، هناك خطر ثالث للبيانات الضخمة؛ حيث إننا قد نقع ضحية استبداد البيانات، الذي بموجبه سنقوم بتقديس المعلومات، ونتائج تحليلاتنا وينتهي بنا الأمر بإساءة استخدامها. عند استخدامها بمسئولية، تعتبر البيانات الضخمة أداة مفيدة تساعد على اتخاذ القرارات المنطقية. أما في حالة استخدامها دون حكمة، فقد تصبح عصا في يد القوى التي قد تحولها إلى مصدر للقمع سواء من خلال تعجيز العملاء والموظفين أو الأسوأ أن يضر بالمواطنين.

إن المخاطر أكبر مما نظن؛ حيث إن مخاطر الفشل في التحكم في البيانات الضخمة فيما يخص الخصوصية والتوقعات، أو تحريف معنى البيانات، يتعدى الأمور التافهة مثل إعلانات الإنترنت التي تستهدف شريحة معينة من المستخدمين. يمتلئ تاريخ القرن العشرين بمواقف دموية ساعدت فيها البيانات على حدوث نتائج قبيحة. في عام 1943، سلم مكتب الإحصاء السكاني الأمريكي عناوين المجمعات السكنية ــ دون أسماء الشوارع وأرقام المنازل، للحفاظ على الوهم الذي يطلقون عليه حماية الخصوصية ـــ التي يسكن بها الأمريكيون اليابانيون من أجل تسهيل اعتقالهم. كما تم استخدام السجلات المدنية الشاملة التي أعدتها هولندا من قبل المحتلين النازيين في تجميع اليهود. كانت الأعداد المكونة من خمسة أرقام التي كانت توشم على أذرع سجناء معسكرات اعتقال النازيين تتوافق مع أرقام البطاقات المثقوبة التي اخترعها "هوليريث" من شركة آي بي إم، لقد سهلت معالجة البيانات من القتل على المستوى الصناعي.

رغم مميزاتها المعلوماتية الكبيرة، كان هناك الكثير من الأمور التي لم تتمكن "ستازي" من القيام بها. لم تكن تتمكن من معرفة تحركات الجميع في جميع الأوقات أو مع من تحدثوا دون بذل جهد جبار. في الوقت الحالي، يتم جمع أغلب هذه المعلومات عن طريق الهواتف المحمولة. لم يكن بمقدور دولة ألمانيا الشرقية أن تتوقع من سينشق عنها، ولا يمكننا نحن أيضًا توقع هذا ـــ ولكن بدأت قوات الشرطة في استخدام نماذج خوارزمية لتقرير متى وأين عليهم أن يقوموا بدورياتهم، الأمر الذي يمنحهم تلميحات بما قد يحدث. وضحت هذه التوجهات أن المخاطر متأصلة في البيانات الضخمة بشكل كبير قد يصل إلى نفس حجم المجموعات البيانية ذاتها.

تعجيز الخصوصية

من المثير أن نحاول استنتاج مخاطر الخصوصية من نمو البيانات الضخمة وأن نرى أمورًا مشابهة لخلل المراقبة الذي ذكره "أوريل" في رواية 1984. ولكن يبدو أن

الموقف أكثر تعقيدًا مما نظن. في البداية، لا تحتوي جميع البيانات الضخمة على بيانات شخصية؛ حيث إن البيانات التي تجمعها الحساسات في مصافي تكرير البترول لا تحتوي على بيانات شخصية، وكذلك لا تحتوي بيانات آلات المصانع أو انفجارات بالوعات الصرف أو بيانات حالة الطقس التي تنشرها المطارات على بيانات شخصية. لا تحتاج (أو ترغب) شركتا بي بي ولا كون إديسون إلى البيانات الشخصية للحصول على القيم المادية من الإحصاءات التي تقومان بها. إن تحليلات البيانات الضخمة من هذا النوع لا تشكل خطرًا عمليًّا على الخصوصية.

ولكن، لا يزال كم كبير من البيانات التي يتم إنتاجها في الوقت الحالي يحتوي على المعلومات الشخصية للأفراد. وتمتلك الشركات دوافع قوية للحصول على المزيد والاحتفاظ بها لوقت أطول وإعادة استخدامها بشكل متكرر. قد لا تبدو البيانات مشابهة للمعلومات الشخصية بشكل جلي، ولكن مع معالجة البيانات الضخمة يمكن تتبع هذه المعلومات وصولاً إلى الأشخاص الذين تشير إليهم. أو يمكن استنتاج بعض التفاصيل الحميمة عن الحياة الشخصية لبعض الأشخاص.

على سبيل المثال، تقوم المؤسسات العامة في الولايات المتحدة الأمريكية وأوربا بإدارة مقاييس كهربية "ذكية" لجمع البيانات على مدار اليوم؛ حيث من المحتمل أن تصل عدد مرات القياس إلى مرة كل 6 ثوان ـــ أكثر بكثير من المعلومات الإجمالية التي جمعتها المقاييس التقليدية عن استهلاك الطاقة. الأمر الرئيسي، هو أن الأجهزة الكهربائية التي تستهلك الطاقة تترك خلفها "بصمة تحميل" تختلف من جهاز لآخر؛ لذا تختلف بصمة غلاية المياه عن الحاسب الآلي، والتي تختلف عن أضواء الصوب الزجاجية. وبالتالي، يكشف استهلاك الطاقة للمنازل عن المعلومات الخاصة به ومن بينها السلوك اليومي لقاطنيه أو الحالة الصحية أو الأنشطة الخارجة على القانون.

السؤال المهم في هذه الحالة ليس ما إذا كانت البيانات الضخمة تزيد من الخطورة على الخصوصية (نعم تفعل)، بل ما إذا كانت تغير من هيئة الخطر ذاته. إذا كان التهديد أكبر بقدر طفيف، فمن المحتمل أن تظل القوانين والقواعد التي تحمي الخصوصية صالحة لعصر البيانات الضخمة، وكل ما سيكون علينا فعله هو مضاعفة جهودنا الحالية. وعلى النقيض، إذا ما تغيرت المشكلة، فسنحتاج إلى حلول جديدة.

لسوء الحظ، لقد تغيرت المشكلة؛ حيث إنه مع البيانات الضخمة، لم تعد قيمة المعلومات تقف عند حدود الغرض الأولي منها فحسب، بل أصبحت في استخداماتها الثانوية، كما ناقشنا من قبل.

لقد قوض هذا التغيير الدور المحوري الذي أُسند إلى الأفراد في قوانين حماية الخصوصية الحالية؛ حيث أصبحت الشركات تخبر الأفراد في وقت جمع معلوماتهم عن المعلومات التي سيتم جمعها ولأي غرض، ومن ثم تتوافر لهم الفرصة للموافقة على ذلك، وبالتالي يمكن أن يبدأ جمع هذه المعلومات. في حين لا يعتبر مبدأ "لاحظ ووافق" هو الطريقة القانونية الوحيدة لجمع البيانات الشخصية ومعالجتها، كما يقول "فريد كايت"، خبير الخصوصية بجامعة إنديانا، فإنها تحولت إلى مبدأ حماية الخصوصية الرئيسي المتبع في جميع أنحاء العالم. (من الناحية العملية، أدى هذا

المبدأ إلى الحصول على كم كبير من ملاحظات حماية الخصوصية التي نادرًا ما يقرؤها المستخدمون، ناهيك عن فهمها ــ ولكن هذا موضوع مختلف).

المثير للدهشة، في عصر البيانات الضخمة، لم تخطر أغلب الاستخدامات الثانوية المبتكرة للبيانات على بال أحد عند جمع هذه البيانات في المقام الأول. كيف يمكن للشركات أن تقدم ملاحظات حماية الخصوصية لأغراض غير موجودة؟ كيف يمكن للأفراد أن يوافقوا على استخدام بياناتهم الشخصية في أمر مجهول؟ ولكن في حالة عدم الموافقة من الأفراد، فإن أي تحليل للبيانات الضخمة يحتوي على بيانات شخصية، يجب أن تعود الشركة إلى كل من هؤلاء الأفراد للحصول على موافقاتهم بإعادة الاستخدام. هل يمكنك تخيل محاولة شركة جوجل التواصل مع مئات الملايين من المستخدمين للحصول على موافقتهم على إعادة استخدام بيانات طلباتهم البحثية على محرك البحث من أجل توقع انتشار الأنفلونزا؟ لا يمكن لأية شركة أن تتحمل مثل هذه التكاليف حتى إن كانت المهمة ممكنة التحقيق من الناحية التقنية.

البديل هو أن يُطلب من العملاء الموافقة على أية استخدامات محتملة لبياناتهم في المستقبل في وقت جمع المعلومات، ولكن هذا الأمر غير مجدٍ هو الآخر؛ حيث إن مثل هذا التصريح الشامل يضعف من فكرة الموافقة عن علم. في سياق البيانات الضخمة، كان مبدأ الملاحظة والموافقة الذي تمت تجربته والثقة به إما محددًا للكشف عن القيمة المستترة في البيانات أو واهيًا فيما يتعلق بحماية خصوصية الأفراد.

هناك طرق أخرى فاشلة لحماية خصوصية الأفراد. إذا كانت المعلومات الشخصية للجميع مجتمعة في مجموعة بيانية واحدة، فإن مجرد اختيار "الانسحاب" من الأمر سيترك أثرًا. خذ مثالاً عن خدمة استعراض الشوارع المقدمة من شركة جوجل؛ حيث تجمع سياراتها صورًا عن الطرق والمنازل في الكثير من بلدان العالم. واجهت جوجل في ألمانيا الكثير من الاعتراضات العامة واعتراضات وسائل الإعلام؛ حيث خشي الناس من أن عرض صور منازلهم وحدائقهم على الإنترنت قد يساعد اللصوص على اختيار الأهداف المربحة. وتحت الضغوط التشريعية، وافقت جوجل على السماح لأصحاب المنازل بالانسحاب من خلال التعتيم على منازلهم في الصور. ولكن يبدو الانسحاب واضحًا في خدمة استعراض الشوارع ــ يمكنك أن تلحظ المنازل التي تم التعتيم عليها ــ وقد يترجم اللصوص هذا الأمر بأنه إشارة على أن هذه المنازل أهداف جيدة لسرقتها.

هناك طريقة تقنية تستخدم لحماية الخصوصية ــ التجهيل ــ والتي لا تعمل بكفاءة في الكثير من الحالات. يعني التجهيل، إزالة أية إشارات تدل على شخصيات الأفراد من المجموعات البيانية، مثل الاسم والعنوان ورقم بطاقة الائتمان وتاريخ الميلاد ورقم التأمين الاجتماعي. ومن ثم يتم تحليل البيانات الناتجة ومشاركتها دون تعريض خصوصية أي شخص للخطر. تصلح هذه الطريقة في عالم البيانات الصغيرة، ولكن البيانات الضخمة مع زيادة كميتها وتنوع المعلومات التي تحتويها، من شأنها تسهيل إعادة تحديد شخصيات الأفراد. تذكر الحالات التي كان من

المفترض أن يكون فيها البحث عبر الإنترنت غير محدد وكذلك تقييم الأفلام.

في أغسطس من عام 2006، أفرجت شركة إيه أو إل عن كمٍّ هائلٍ من طلبات البحث القديمة، بموجب وجهة النظر السليمة بأن الباحثين قادرون على تحليلها للحصول على أفكار شيقة. كان قد تم تجهيل هذه المجموعة البيانية بعناية، التي تكونت من 20 مليون طلب بحثي من 657 ألف مستخدم في الفترة ما بين 31 مارس و31 مايو من العام ذاته. كان قد تم محو أسماء المستخدمين وعناوين بروتوكولات الإنترنت الخاصة بهم ووضع مؤشرات عددية بدلاً منها. كانت الفكرة تدور حول قدرة الباحثين على الربط بين طلبات البحث الخاص بالشخص ذاته، ولكن دون وجود أية معلومات تدلهم على شخصيات المستخدمين الحقيقية.

ولكن، بعد عدة أيام، قامت جريدة نيويورك تايمز بالمزج بين طلبات بحث على غرار "60 رجلاً أعزب" و"الشاي المفيد للصحة" و"المناظر الطبيعية في ليلبرن بجورجيا" لتتمكن من تحديد هوية المستخدم رقم 4417749 على أنها "تيلما أرنولد" الأرملة البالغة من العمر 62 عامًا من ليلبرن جورجيا، والتي قالت لمراسل الجريدة عندما ذهب لزيارتها قائلة: "يا إلهي، إنها حياتي الشخصية كاملة. لم أكن أعلم أن هناك من يتجسس عليّ". وتسببت الاحتجاجات العامة المترتبة على هذا الأمر في طرد مدير القسم التقني بالشركة واثنين من الموظفين.

ولكن بعد مرور شهرين، في شهر أكتوبر عام 2006، قامت خدمة إيجارات الأفلام "نتفلكس" بأمر مشابه عندما أطلقت مسابقة "جائزة نتفلكس"؛ حيث أفرجت الشركة عن 100 مليون سجل لإيجارات الأفلام من حوالي نصف مليون مستخدم ـــ وعرضت جائزة مليون دولار لأي فريق يتمكن من تحسين نظامها لتوصيات الأفلام بمقدار 10 %. وقامت هي الأخرى بإزالة أية دلالات شخصية من البيانات، ومرة أخرى تم تحديد هوية أحد المستخدمين: وكانت أمًّا تم تحديد هويتها في الوسط الغربي المحافظ، والتي قامت بعد ذلك برفع دعوى قضائية على الشركة مستخدمة اسمًا مستعارًا هو "جاين دو".

قام الباحثون بجامعة تكساس بولاية أوستن بمقارنة بيانات نتفلكس مع غيرها من المعلومات العامة. وسرعان ما اكتشفوا أن التقييمات التي قام بها المستخدمون المجهولون توافقت مع المشاركين المعروفين في موقع قاعدة بيانات الأفلام على الإنترنت. وأثبت البحث أن تقييم ستة أفلام مجهولة (خارج نطاق الخمسمائة فيلم الأولى في التصنيف) من شأنه المساعدة على تحديد هوية أحد عملاء نتفلكس بنسبة 84% في كل مرة. وإذا كان المرء يعرف التاريخ الذي قام فيه المستخدم بتقييم الفيلم، فمن الممكن أن يتم تحديد هويته من بين النصف مليون مستخدم بدقة قد تصل إلى 99%.

في حالة إيه أو إل، تعرضت هويات المستخدمين للخطر عن طريق محتوى كلمات البحث التي كتبوها في محركات البحث. في حالة نتفلكس، تم تحديد هوية المستخدمين من خلال مقارنة البيانات مع مصادر أخرى. في كلتا الحالتين، لم تدرك الشركتان أن البيانات الضخمة تساعد على عدم القدرة على تجهيل البيانات. هناك سببان لهذا الأمر: أننا نجمع المزيد من البيانات وأننا نمزج بين المزيد من

المجموعات البيانية.

يوضح "بول أوم"، أستاذ القانون بجامعة كولورادو ببولدر والخبير في الأضرار التي يسببها عدم تجهيل البيانات، أنه لا يوجد حل سهل متوافر لهذا الأمر؛ ففي ظل وجود وفرة في البيانات، فإن التجهيل يكون مستحيلاً بغض النظر عن كم المحاولات التي يقوم بها المرء. الأمر الأسوأ أن الباحثين قد اكتشفوا مؤخرًا أنه ليست البيانات التقليدية فقط هي المعرضة لاستحالة التجهيل، بل المخطط الاجتماعي أيضًا ـــ علاقات الآخرين بعضهم بالآخر.

في عصر البيانات الضخمة، فقدت الإستراتيجيات الثلاث الرئيسية التي طالما تم استخدامها للتأكيد على حماية الخصوصية ـــ الملاحظة والموافقة للأفراد، والانسحاب، والتجهيل ـــ الكثير من فاعليتها؛ حيث يشعر الكثير من المستخدمين في الوقت الحالي بأن خصوصيتهم تنتهك. انتظر فقط حتى تصبح ممارسات البيانات الضخمة أكثر انتشارًا.

بالمقارنة مع دولة ألمانيا الشرقية منذ ربع قرن مضى، أصبحت المراقبة أكثر سهولة وأقل تكلفة وأكثر قوة؛ حيث إن القدرة على الحصول على البيانات الشخصية أصبحت تدمج بشكل عميق في الأدوات التي نستخدمها يوميًّا، من المواقع الإلكترونية إلى تطبيقات الهواتف الذكية. إن مسجلات البيانات المتواجدة في أغلب السيارات والتي تعمل على جمع جميع ما يحدث في السيارة قبل لحظات من انفتاح الوسائد الهوائية بها، تُستخدم "كشاهد" ضد مالكي السيارات خلال الخصومات القضائية المتعلقة بأسباب الحوادث.

لا شك في أنه عندما تقوم الشركات بجمع البيانات من أجل تحسين أرباحها، لا حاجة بنا لأن نخاف أن مراقبتهم لنا ستحمل نفس عواقب تجسس "ستازي" على مواطني ألمانيا الشرقية؛ حيث إننا لن نوضع بالسجن إذا ما اكتشفت أمازون أننا نحب قراءة كتاب "ما تسي تونج" تحت عنوان "الكتاب الأحمر الصغير"، ولن تقوم جوجل بنفينا بسبب بحثنا عن كلمة "بنج". قد تتمتع هذه الشركات بقوة كبيرة، ولكنها لا تملك السلطة لإجبارنا على أي شيء.

لذا، رغم أنها لا تستطيع أن تلقي القبض علينا في منتصف الليل، فإنها تكدس جبالاً من بياناتنا الشخصية المتعلقة بجميع جوانب حياتنا، وتشاركها مع الآخرين دون معرفتنا، وتستخدمها بطرق لا يمكننا تخيلها.

لا يستعرض القطاع الخاص وحده قوته في مجال البيانات الضخمة؛ حيث إن الحكومات تقوم بالأمر ذاته. على سبيل المثال، وكالة الأمن القومي الأمريكي يقال إنها تعترض وتخزن 1.7 مليار من الرسائل الإلكترونية والمحادثات الهاتفية وغيرها من طرق التواصل كل يوم، طبقًا للتحقيق الذي أجرته جريدة واشنطن بوست عام 2. يقدر "ويليام بيني"، المسئول السابق بوكالة الأمن القومي الأمريكي، أن الحكومة قد جمعت حوالي "20 تريليون معاملة" بين المواطنين الأمريكيين وغيرهم ـــ من يتصل بمن، ويراسل من إلكترونيًّا ويرسل الأموال لمن، وغيرها من الأمور.

من أجل فهم جميع هذه البيانات، تقوم الولايات المتحدة الأمريكية ببناء مراكز

عملاقة للبيانات مثل منشأة الأمن القومي الأمريكي التي تكلفت 1.2 مليار دولار في فورت ويليامز بولاية يوتاه. وتطالب جميع المصالح الحكومية كميات من المعلومات أكثر مما سبق، وليس فقط الوكالات السرية التي تعمل على مقاومة الإرهاب. عندما يتوسع جمع البيانات ليشمل التعاملات المالية والسجلات الصحية وتحديثات الحالات على موقع فيس بوك، فإن كم البيانات التي يتم جمعها يكون هائلاً، ولا يمكن للحكومة أن تقوم بتحليل هذا الكم الهائل من البيانات، فلماذا إذن تقوم بجمعها؟

لقد تطورت أساليب المراقبة في عصر البيانات الضخمة؛ ففي الماضي كان المحققون يستخدمون المشابك التي تشبه فكوك التماسيح لتثبيتها على خطوط الهاتف ومعرفة المزيد عن المشتبه بهم؛ حيث كان المهم حينها هو التحرك السريع من أجل معرفة المزيد عن هذا الشخص. أما الأسلوب الحديث فيختلف تمام الاختلاف؛ فمن وجهة نظر جوجل أو فيس بوك، يدور التفكير الحديث حول أن البشر هم مجموع علاقاتهم الشخصية وتفاعلاتهم على شبكة الإنترنت وتواصلهم مع محتواها. ومن أجل تحقيق المعرفة الكاملة عن شخص ما، على المحللين أن يلقوا نظرهم على أكبر كم ممكن من هالة البيانات التي تحيط بالشخص ــ ليس فقط الأشخاص الذين يعرفهم، بل أيضًا الأشخاص الذين يعرفهم هؤلاء الأشخاص، وهكذا. كان هذا الأمر صعبًا للغاية من الناحية التقنية في الماضي. أما اليوم، فقد أصبح هذا الأمر أسهل من أي وقت مضى. ولأن الحكومة لم تتمكن قط من معرفة من عليها أن تراقب، فإنها تقوم بجمع وتخزين أو ضمان إمكانية الوصول إلى المعلومات ليس من أجل مراقبة الجميع طوال الوقت بالضرورة، ولكن عندما يقع شخص ما داخل دائرة الاشتباه، يمكن للسلطات على الفور أن تتحرى عنه بدلاً من بدء جمع المعلومات عنه من جديد.

إن الولايات المتحدة الأمريكية ليست الحكومة الوحيدة التي تجمع جبالاً من البيانات عن شعوبها، وليست أكثر الحكومات التي تستخدم الوسائل السافرة للقيام بهذا الأمر. ولكن، الأمر الذي يمثل مشكلة تعادل في أهميتها مشكلة قدرة الشركات والحكومات على معرفة معلوماتنا الشخصية، هو أمر جديد ظهر مع ظهور البيانات الضخمة: استخدام التوقعات في الحكم علينا.

الاحتمالية والعقاب

"جون أندرتون"، رئيس واحدة من وحدات الشرطة الخاصة في العاصمة الأمريكية واشنطن. في صباح أحد الأيام، تمكن من اقتحام أحد المنازل في ضواحي المدينة قبل لحظات من طعن "هوارد مارك" ـــ وهو في حالة شديدة من الغضب - زوجته بمقص في بطنها بعدما وجدها تخونه مع شخص آخر. بالنسبة "لأندرتون"، إن هذا هو المعتاد في عمله لمنع الجرائم الكبيرة؛ حيث قال: "نيابة عن قسم استباق الجرائم بمقاطعة كولومبيا، ألقي القبض عليك لارتكابك جريمة القتل في المستقبل بحق سارة مارك، والتي كانت ستقع اليوم...".

قام بعض رجال الشرطة الآخرون باعتقال "مارك" الذي صرخ قائلاً: "أنا لم أفعل

أي شيء".

يصور المشهد الافتتاحي من فيلم *Minority Report* مجتمعًا تبدو فيه التوقعات على قدر كبير من الدقة لدرجة أن الشرطة تلقي القبض على الناس لارتكابهم جرائم قبل حتى أن يرتكبوها. ويتم إيداعهم في السجن، ليس على جرائم ارتكبوها، بل على جرائم تم التوقع بارتكابهم إياها، رغم أنهم لم يرتكبوا الجريمة بالفعل. يسند الفيلم هذا التطبيق القانوني المستقبلي والوقائي إلى رؤى ثلاثة من المتبصرين، وليس إلى تحليل البيانات. إن المستقبل المضطرب الذي يصوره فيلم مينيوريتي ريبورت من بين الأمور التي يهدد تحليل البيانات الضخمة الطليق بأن يحققه؛ حيث ستكون الأحكام في هذا العصر قائمة على توقعات السلوكيات المستقبلية للأفراد.

لقد رأينا بوادر هذا المستقبل تحدث بالفعل. تستخدم لجان إطلاق السراح المشروط في أكثر من نصف ولايات الولايات المتحدة الأمريكية التوقعات القائمة على تحليل البيانات كعامل مساعد على اتخاذ القرار بإطلاق سراح أحد السجناء أم إبقائه حبيسًا. هناك أعداد متزايدة من الأماكن في الولايات المتحدة الأمريكية ــ من بعض أقسام الشرطة في لوس أنجلوس إلى مدن مثل ريتشموند بولاية فيرجينيا ــ تستخدم "العمل الشرطي التوقعي": استخدام تحليل البيانات الضخمة لاختيار الشوارع والمجموعات والأفراد الذين يجب أن يخضعوا للمزيد من الترتيبات الأمنية؛ لأن المعادلات الخوارزمية قد أشارت إليهم متوقعة أنهم قد يرتكبون جرائم.

في مدينة ممفيس بولاية تينيسي، هناك برنامج يدعى بلو كراش (برنامج الحد من الجرائم باستخدام التاريخ الإحصائي) يقدم لضباط الشرطة مناطق محددة نسبيًّا للاهتمام بها من حيث الموقع (أحياء قليلة) والوقت (خلال ساعات قليلة من يوم معين من أيام الأسبوع). يساعد النظام بشكل واضح على تطبيق القانون بشكل أفضل من خلال مساعدة الشرطة على تحديد مواردها النادرة. منذ بداية العمل به عام 2006، قلت نسبة الجرائم الكبرى ضد الملكيات وجرائم العنف إلى الربع، طبقًا لإحدى الإحصائيات (ولكن لا يوجد أي دليل على سبب هذا الانخفاض؛ حيث لا يوجد ما يدل على أن هذا البرنامج هو ما سبب انخفاض تلك النسبة).

في مدينة ريتشموند بولاية فيرجينيا، يربط رجال الشرطة بين بيانات الجرائم ومجموعات بيانية إضافية، مثل معلومات عن أماكن دفع الشركات الكبيرة في المدينة لرواتب موظفيها أو تواريخ الحفلات أو الأحداث الرياضية. كان لهذا الأمر تأثير إيجابي على تأكيد من يشتبه بهم رجال الشرطة وأحيانًا عدلت من اشتباههم بالنزعات الإجرامية لدى بعض الناس. على سبيل المثال، كان رجال الشرطة في ريتشموند يشعرون دائمًا بزيادة معدلات الجرائم بعد عروض الأسلحة، أثبت تحليل البيانات الضخمة أنهم على حق ولكن مع ملحوظة صغيرة: أن الارتفاع قد زاد بعد أسبوعين من العرض وليس بعد العرض مباشرة.

تسعى هذه الأنظمة إلى منع الجرائم من خلال التوقع، وصولاً في نهاية الأمر إلى مستوى الأفراد الذين قد يرتكبونها. يشير هذا الأمر إلى استخدام البيانات الضخمة لغرض جديد: منع الجرائم قبل حدوثها.

حاول مشروع أبحاث تابع لإدارة الأمن القومي الأمريكية يُطلق عليه فاست (تقنية فحص صفات المستقبل) أن يحدد الإرهابيين المحتملين من خلال مراقبة العلامات الحيوية للأفراد ولغة جسدهم وغيرها من الأنماط النفسية. تدور فكرة المشروع حول أن مراقبة سلوك الناس قد يكشف عن نيتهم للقيام بالأذى. خلال اختباره، بلغت دقة النظام 70 %، طبقًا لإدارة الأمن القومي. (إن ما يعنيه هذا غير واضح، هل كان الأشخاص الخاضعون للبحث يتظاهرون بأنهم إرهابيون ليتم تحديد "نياتهم السيئة"؟) رغم أن هذه الأنظمة لم تنضج بعد، فإن الأمر المهم هو أن سلطات تطبيق القانون تأخذها على محمل الجد.

يبدو أن منع الجريمة قبل حدوثها احتمال مُغْرٍ. ألن يكون منع الجرائم قبل حدوثها أفضل من عقاب المجرمين بعد حدوثها؟ ألن يفيد إدراك الجرائم بشكل مسبق ليس فقط ضحايا تلك الجرائم بل المجتمع ككل؟

ولكن إن هذا الطريق محفوف بالمخاطر. إذا تمكنا من خلال تحليل البيانات الضخمة من أن نتوقع من قد يرتكب جرائم في المستقبل، فقد لا نرتاح إلى مجرد منع حدوث تلك الجرائم؛ حيث إننا من المحتمل أن نرغب في عقاب المجرمين المحتملين أيضًا. إنه أمر منطقي. إذا ما تدخلنا لمنع حدوث الفعل المخالف للقانون، فإن المجرم المحتمل قد يحاول أن يفلت دون عقاب مرة أخرى. ولكن على النقيض، من خلال استخدام البيانات الضخمة لتحميله مسئولية أفعاله (المستقبلية)، فقد نتمكن من ردعه ومن ردع آخرين غيره.

يبدو أن أنواع العقاب تلك التي تعتمد على التوقعات قد أصبحت تحسينًا للممارسات التي بدأنا في قبولها؛ حيث إن منع السلوكيات الضارة أو الخطرة هو حجر الأساس للمجتمعات الحديثة. لقد جعلنا من عادة التدخين أمرًا صعبًا للوقاية من سرطان الرئة؛ وفرضنا ربط أحزمة الأمان لتجنب الوفيات في حوادث السيارات؛ وأصبحنا نمنع صعود المسافرين إلى الطائرات وهم يحملون أسلحة لتجنب اختطاف الطائرات. قد تقيد هذه الإجراءات الوقائية من حريتنا، ولكن يراها الكثيرون على أنها ثمن بخس في مقابل تجنب الضرر الأكبر.

في الكثير من الحالات، تم بالفعل توظيف البيانات تحت مسمى الوقاية؛حيث تم استخدامها لجمعنا داخل جماعات من الأشخاص الذين يشبهوننا، وعادة ما يتم تصنيفنا طبقًا لهذه الجماعات. لاحظت الجداول الاكتوارية أن الرجال الذين تزيد أعمارهم على 50 عامًا يكونون أكثر عرضة لسرطان البروستاتا؛ لذا ربما يقوم أعضاء هذه المجموعة بدفع المزيد من تكاليف التأمين الصحي حتى إن لم يصابوا بسرطان البروستاتا. عادة ما يكون طلبة المدارس الثانوية ذوو الدراجات العالية، كمجموعة، أقل عرضة لحوادث السيارات؛ لذا يدفع بعض من أقرانهم ذوي الدراجات الأقل أقساطًا تأمينية أعلى. يخضع الأشخاص ذوو السمات الخاصة للمزيد من التفتيش في أثناء مرورهم بأمن المطارات.

هذه هي الفكرة التي تكمن خلف "السمات العامة" في عالم البيانات الصغيرة المعاصر. حدد سمات مشتركة داخل البيانات، وحدد مجموعة من الأشخاص تنطبق عليهم هذه السمات، ومن ثم ضع هؤلاء الأشخاص تحت منظار الأمن بشكل أكبر

من غيرهم، وهي قاعدة تنطبق على جميع أعضاء هذه المجموعة. لا شك في أن عبارة "السمات العامة" تحتوي على الكثير من المعاني، وأن الأسلوب له الكثير من الجوانب السيئة. إذا ما أُسيء استخدام هذا الأسلوب، فإنه لن يؤدي إلى التحامل في معاملة مجموعة بعينها فحسب، بل قد يؤدي أيضًا إلى "الاتهام لمجرد الانتماء لمجموعة ما".

وعلى النقيض، تختلف توقعات البيانات الضخمة فيما يتعلق بالأشخاص عما سبق؛ حيث تعتمد التوقعات اليومية عن السلوكيات المحتملة ــ التي نجدها في أمور مثل أقساط التأمين أو التقارير الائتمانية ــ على عدد صغير من العوامل التي تعتمد بدورها على النماذج العقلية للمشكلة التي بين أيدينا (أي، المشكلات الصحية السابقة أو تاريخ رد القروض)، من خلال التحليل المحايد للبيانات الضخمة، عادة ما نتمكن من تحديد أكثر التوقعات الملائمة داخل هذا الكم الهائل من المعلومات.

الأمر الأهم، أننا نستخدم البيانات الضخمة على أمل منا بأن نتمكن من تحديد أشخاص بعينهم بدلاً من المجموعات؛ حيث سيحررنا هذا الأمر من عيوب السمات العامة التي تعرض كل توقع لشبهة أن يكون حالة من حالات الاتهام لمجرد الانتماء للمجموعة. في عالم البيانات الضخمة، لن يخضع الشخص العربي، الذي يحجز تذكرة في اتجاه واحد في الدرجة الأولى على إحدى الرحلات الجوية، إلى المزيد من التفتيش في المطار إذا ما قالت بياناته إنه من غير المحتمل أن يكون إرهابيًّا. باستخدام البيانات الضخمة يمكننا أن نهرب من السمات العامة التي تميز هويات مجموعات الأشخاص، ونستبدل بها توقعات مدققة أكثر لكل فرد على حدة.

تعدنا البيانات الضخمة بأننا سنظل نقوم بما اعتدنا القيام به طوال الوقت ــ السمات العامة ــ ولكنها ستقوم بتحسينها وتقلل من التمييز الذي تتضمنه، وتركز أكثر على الأفراد. يبدو هذا الأمر مقبولاً إذا ما كان الهدف هو منع الأفعال غير المرغوب بها. ولكن الخطورة هي ما إذا استخدمنا توقعات البيانات الضخمة لتقرير ما إذا كان الشخص مذنبًا وأنه تجب معاقبته على السلوك الذي لم يقم به بعد.

إن فكرة العقاب بناءً على المجرمين وليس على الجرائم مقززة؛ حيث إن اتهام شخص ما بناء على سلوكه المستقبلي المحتمل يلغي العدالة من أساسها: حيث إن المرء يجب أن يقوم بفعل ما قبل أن نحمله مسئوليته. مع ذلك، فإن التفكير في أفكار سيئة ليس مخالفًا للقانون، بل إن القيام بالأمور السيئة هو الأمر المخالف للقانون. إن العقيدة الرئيسية في مجتمعنا هي أن مسئولية الفرد ترتبط بخياراته. إذا ما أُجبر شخص ما على فتح خزينة الشركة تحت تهديد السلاح، فإن هذا الشخص لم يكن له خيار سوى فعل ذلك، ولا يعتبر مسئولاً عن هذا الفعل.

إذا كانت توقعات البيانات الضخمة دقيقة، وإذا ما تمكنت الخوارزميات من توقع مستقبلنا بوضوح تام، فلن يكون لنا خيار فيما نفعله في المستقبل؛ حيث كنا سنتصرف بالضبط طبقًا للمتوقع. إذا ما كانت التوقعات الدقيقة محتملة، فستسلبنا إرادتنا، قدرتنا على عيش حياتنا بحرية. وكذلك، من خلال حرماننا من حرية الاختيار، فإنها تزيل من على عاتقنا أية مسئولية.

لا شك في أن التوقعات الدقيقة أمر مستحيل الحدوث، ولكن سيتوقع تحليل

البيانات الضخمة أنه بالنسبة لشخص ما، سيكون للسلوك المستقبلي الذي سيقوم به احتمال محدد. تأمل، على سبيل المثال، في البحث الذي أجراه "ريتشارد بيرك"، أستاذ الإحصاء وعلم الجريمة بجامعة بنسلفانيا، والذي افترض أن أسلوبه يمكنه توقع ما إذا كان الشخص الذي خرج من السجن ضمن إطلاق السراح المشروط سيتورط في جريمة قتل (سواء كان قاتلاً أو مقتولاً). تعددت المدخلات التي استخدمها في البحث والتي استمدها من متغيرات الكثير من القضايا، بما فيها سبب الاعتقال وتاريخ أول جريمة ارتكبها، وكذلك البيانات الديموغرافية مثل العمر والنوع. يقترح "بيرك" أنه قادر على توقع جرائم القتل المستقبلية التي من المحتمل أن يتورط فيها المجرمون الذين أطلق سراحهم طبقًا لبرنامج إطلاق السراح المشروط بنسبة احتمالية بلغت 70 % على أقل تقدير. وهذه النسبة ليست سيئة، ولكنها تعني أيضًا أنه على لجان إطلاق السراح المشروط أن تعتمد على تحليل " بيرك"؛ حيث إن نسبة خطئهم عادة ما تكون واحدًا لكل 4 حالات.

ولكن المشكلة الجوهرية المتعلقة بالاعتماد على هذه التوقعات ليس تعريض المجتمع للخطر. المشكلة الرئيسية أنه باستخدام هذا النظام سنعاقب الناس قبل أن يرتكبوا فعلاً سيئًا. ومن خلال تدخلنا قبل أن يقوموا بهذه الأفعال السيئة (على سبيل المثال، حرمانهم من إطلاق السراح المشروط إذا ما أظهرت التوقعات وجود احتمال كبير لأنهم قد يتورطون في جرائم قتل)، فإننا لن نعلم أبدًا ما إذا كانوا سيرتكبون هذه الجرائم بالفعل أم لا. إننا لا نترك الأحداث لتقوم بدورها، ولكننا نحمل الأشخاص مسئولية توقعاتنا لما سوف يقومون بفعله. ولا يمكن نفي هذه التوقعات على الإطلاق.

يلغي هذا الأمر فكرة افتراض البراءة، المبدأ الذي يقوم عليه نظامنا القانوني وإحساسنا بالعدالة. إذا ما حملنا الناس مسئولية أفعالهم المستقبلية المتوقعة، الأفعال التي قد لا يقومون بها أبدًا، فإننا ننكر أيضًا أن البشر قادرون على اختيار خيارات أخلاقية.

إن النقطة المهمة في هذا الموضوع، ليست إنفاذ القانون؛ حيث إن الخطر أكبر بكثير من مجرد تقديم المجرمين للعدالة؛ حيث إنها تغطي جميع أوجه المجتمع، جميع حالات الحكم البشري التي يتم استخدام البيانات الضخمة فيها لتقرير ما إذا كان الناس يتحملون مسئولية أفعالهم المستقبلية أم لا. يشمل هذا الأمر كل شيء من اتخاذ الشركة القرار بطرد موظف ما من العمل، إلى رفض الطبيب إجراء عملية جراحية لأحد المرضى، إلى رفع إحدى الزوجات دعوى طلاق على زوجها.

ربما ساعدت هذه الأنظمة على جعل المجتمع أكثر أمانًا وفاعلية، ولكن الجزء الرئيسي الذي يجعلنا من البشر ــ قدرتنا على اختيار أفعالنا وتحمل مسئوليتها ــ ربما يكون قد تحطم. ربما أصبحت البيانات الضخمة أداة لجعل الاختيارات البشرية تعمل وفقًا لمبادئ المجتمع وتمحو الإرادة الحرة منه.

لا شك في أن البيانات الضخمة تقدم الكثير من الفوائد، ولكن ما حوَّلها إلى سلاح لسلب إنسانيتنا هي العيوب، ليس عيوب البيانات الضخمة ذاتها، بل الطريقة التي تُستخدم بها توقعاتها. إن أساس الموضوع هو تحميل الناس مسئولية الأفعال

المتوقعة قبل أن يقوموا بها باستخدام توقعات البيانات الضخمة بناءً على العلاقات التبادلية لاتخاذ قرارات مسببة عن المسئولية الفردية.

تفيد البيانات الضخمة في فهم المخاطر الحالية والمستقبلية، ولأن نعدل من أفعالنا طبقًا لذلك؛ حيث تساعد توقعاتها المرضى وشركات التأمين والمقرضين والمستهلكين. ولكن لا تخبرنا البيانات الضخمة بكل شيء عن الأسباب. ولكن على النقيض من هذا، يتطلب تحديد "الذنب" ــ المسئولية الفردية ــ أن يختار الأشخاص الذين نحكم عليهم بالقيام بفعل معين. يجب أن يكون قرارهم سببًا في حدوث الفعل الذي يتبعه. وحيث إن البيانات الضخمة تعتمد على العلاقات التبادلية، فإنها لا تصلح كأداة لمساعدتنا على الحكم على الأسباب وبالتالي تحديد المسئولية الفردية.

المشكلة أن البشر قد شبوا على رؤية العالم من منظار السبب والتأثير؛ لذا فإن البيانات الضخمة تقع تحت التهديد الدائم لإساءة استخدامها للأغراض السببية، أن نكون متقيدين بالرؤى الوردية عن مدى فعالية أحكامنا، واتخاذنا القرارات البشرية المتعلقة بتحديد المسئولية، لن تتحقق إلا إذا كنا مسلحين بتوقعات البيانات الضخمة.

هذا هو الطريق الرئيسي المحفوف بالمخاطر ــ الذي يقودنا مباشرة إلى المجتمع الذي تم تصويره في فيلم *Minority Report*، العالم الذي تم محو الخيار الفردي والإرادة الحرة منه، الذي تم فيه استبدال بوصلتنا الأخلاقية الفردية ووضع الخوارزميات التوقعية بدلًا منها والذي يتعرض فيه الأفراد إلى تعسف الأوامر الجماعية. إذا ما حدث هذا الأمر، فإن البيانات الضخمة ستسجننا ــ ربما حرفيًّا ــ في الاحتمالات.

ديكتاتورية البيانات

قامت البيانات الضخمة بإضعاف حماية الخصوصية وتهديد الحرية، إنها تفاقم أيضًا من مشكلة قديمة: الاعتماد على الأرقام رغم أنها معرضة لأن تكون خاطئة أكثر مما نظن. لا يوجد ما يؤكد على عواقب انحراف تحليل البيانات عن مساره أفضل من قصة "روبرت ماكنامارا".

كان "ماكنامارا" شخصًا يعتمد بشكل كامل على الأرقام. وكان قد عُين وزيرًا للدفاع في حكومة الولايات المتحدة الأمريكية في أثناء بدء المعارك في فيتنام في ستينيات القرن العشرين. كان "ماكنامارا" يصر بشدة على الحصول على جميع البيانات التي يمكن الحصول عليها. كان يعتقد أنه من خلال تطبيق الإحصائيات الدقيقة، يمكن لصناع القرار أن يستوعبوا المواقف المعقدة وأن يتخذوا القرارات الصحيحة. كان العالم في نظره كم كبير من المعلومات الجامحة التي إذا ما تم ترتيبها والعناية بها وتحديد الخطوط العريضة لها وحددت كمياتها سيمكن استئناسها بواسطة البشر وقد تخضع لإرادة البشر. كان "ماكنامارا" يسعى خلف الحقيقة، ويمكن إيجاد هذه الحقيقة في البيانات. من بين الأرقام التي وصلت إليه "عدد الجثث".

طور "ماكنامارا" عشقه للأرقام منذ أن كان طالبًا بكلية إدارة الأعمال بجامعة هارفارد ثم عمله بها كأصغر أستاذ مساعد في تاريخها وهو في الرابعة والعشرين من عمره. لقد طبق هذا المبدأ خلال الحرب العالمية الثانية كجزء من فريق صفوة البنتاجون الذي كان يُطلق عليه الرقابة الإحصائية، والذي كان يقدم قرارات مبنية على البيانات لواحدة من أكبر المؤسسات الحكومية في العالم. قبل هذا كانت القوات المسلحة الأمريكية عمياء؛ حيث لم تكن تعلم، على سبيل المثال، نوع أو كمية أو مكان الحصول على قطع غيار الطائرات. ولكن كانت البيانات متواجدة ومتأهبة للإنقاذ. فمن خلال جعل تدابير التسليح أكثر فاعلية تسببت في توفير 3.6 مليار دولار عام 1943؛ حيث كانت الحرب الحديثة تعتمد على التحديد الفعال للموارد؛ وقد حقق الفريق نجاحًا باهرًا.

بعد نهاية الحرب، قرر أعضاء الفريق استمرارهم في العمل معًا وتقديم مهاراتهم للمؤسسات الأمريكية. كانت شركة فورد للسيارات تتخبط، وسلمهم "هنري فورد الثاني" اليائس زمام الأمور. ومثلما لم يكونوا يعرفون أي شيء عن الحرب عندما ساعدوا على الانتصار فيها، لم يكونوا يفقهون شيئًا في صناعة السيارات. ولكن، قام من كانوا يطلقون على أنفسهم "الأطفال البارعون" بقلب حال الشركة رأسًا على عقب.

مر "ماكنامارا" سريعًا على جميع ما تصنعه الشركة، مستخرجًا البيانات عن جميع المواقف. أنتج مديرو المصانع المتعجلون النماذج التي طلبها ــ سواء كانت صحيحة أم لا. عندما أصدر قرارًا بأن جميع مخزون الشركة من أحد طرز السيارات يجب أن يتم استخدامه قبل البدء بإنتاج الطراز الجديد، قام مديرو خطوط الإنتاج الساخطون بإلقاء قطع الغيار الزائدة عن حاجتهم في النهر المجاور. وأومأ المديرون الكبار في المقر الرئيسي بالموافقة عندما أرسل رؤساء العمال إليهم الأرقام الدالة على أنهم قد أطاعوا الأمر. ولكن كانت الدعابة المتداولة في المصنع هي أن المرء قادر على السير على الماء ــ فوق القطع الصدئة من سيارات عام 1950 و1951.

لقد لخص "ماكنامارا" ما يجب أن يكون عليه مديرو منتصف القرن العشرين، المدير التنفيذي الذي يفكر بمنطقية شديدة والذي يعتمد على الأرقام بدلاً من المشاعر، والذي كان قادرًا على تطبيق مهاراته الكمية على أي مجال يعمل به. في ستينيات القرن العشرين، أصبح رئيس شركة فورد، المنصب الذي شغله لأسابيع قليلة قبل أن يعينه الرئيس "كينيدي" وزيرًا للدفاع.

مع تزايد حدة الصراع في فيتنام، وإرسال الولايات المتحدة الأمريكية للمزيد من القوات إلى هناك، أصبح جليًّا أن هذه الحرب هي حرب من أجل فرض الإرادة وليست حربًا من أجل الأرض. كانت خطة الأمريكان تهدف إلى إرغام الفيت كونج على اللجوء لطاولة المفاوضات، وكانت طريقة قياس التقدم هي أعداد القتلى من جيش العدو، وكان يتم نشر عدد الجثث يوميًّا في الصحف. كان هذا الأمر بالنسبة لمناصري الحرب دليلاً على إحراز تقدم، أما بالنسبة للمناهضين لها، فإنه دليل على انعدام أخلاقيتها. كان عدد القتلى هي السمة البيانية الأهم في تلك الحقبة.

في عام 1977، بعد رحيل آخر مروحية عن سقف السفارة الأمريكية في سايجون، نشر أحد الجنرالات المتقاعدين ويُدعى "دوجلاس كينارد" تحقيقًا مميزًا عن وجهات نظر الجنرالات عن هذه الحرب. كشف الكتاب، الذي كان عنوانه "مديرو الحرب"، عن كم هائل من الإحصاءات؛ حيث اعتبرت نسبة 2 % فقط من جنرالات أمريكا أن عدد القتلى كان طريقة مناسبة لقياس التقدم، في حين قال حوالي ثلثي الجنرالات إن هذه الأرقام كانت مبالغة. كتب أحد الجنرالات الآخرين: "كانت عادة عبارة عن أكاذيب وقحة"، وقال آخر: "كانت تتم المبالغة في هذه الأرقام بواسطة وحدات القتال بشكل أساسي بسبب الاهتمام الذي أظهره الكثير من الأشخاص مثل ماكنامارا".

مثلما فعل العاملون في مصنع شركة فورد عندما ألقوا قطع غيار المحركات في النهر، كان صغار الضباط يقدمون أرقامًا مبالغًا فيها إلى قادتهم من أجل استمرارهم في مواقعهم أو إعطاء دفعة لمسيرتهم المهنية؛ حيث أخبروا قادتهم بما يريدون سماعه. كان "ماكنامارا" ومن يحيطون به يعتمدون على الأرقام، ويهتمون بها كثيرًا. شعر "ماكنامارا"، بشعره المصفف للخلف بعناية وربطة عنقه الخالية من العيوب، بأنه لن يتمكن من استيعاب ما يحدث على الأرض إلا من خلال مراجعة الجداول الممتدة ــ جميع هذه الصفوف والأعمدة المرتبة، والحسابات والرسومات البيانية، التي بدا أن براعته فيها تجعله يكاد يكون لا يخطئ.

إن الاستخدام وإساءة الاستخدام والاستخدام الخاطئ للبيانات من قبل القوات المسلحة الأمريكية خلال حرب فيتنام درس قاسٍ عن حدود المعلومات في عصر البيانات الصغيرة، درس يجب الانتباه له جيدًا في أثناء هرولة العالم مسرعًا نحو حقبة البيانات الضخمة. يمكن أن تكون جودة البيانات الأساسية ضعيفة، وقد تكون متحيزة، ويمكن أن تُحلل أو تُستخدم بشكل خاطئ. الأمر الأخطر أن البيانات قد تفشل في التوصل إلى ما تهدف إلى قياسه.

إننا معرضون إلى "ديكتاتورية البيانات" أكثر مما نعتقد ــ أي ترك البيانات لتحكم علينا بطرق قد تسبب الضرر كما تسبب الفائدة. يكمن الخطر في أننا سندع أنفسنا مقيدين دون وعي إلى نتائج تحليلاتنا حتى إن بدت أمامنا شكوك واضحة بأن هناك أمرًا ما خاطئًا. أو أننا سنصبح مهووسين بجمع الحقائق والأرقام من أجل البيانات. أو أننا سنسند للبيانات درجة من الصدق لا تستحقها.

هناك الكثير من أوجه حياتنا قد تم تحويلها إلى بيانات؛ حيث إن الحل الذي بدأ صناع القرار ورجال الأعمال في محاولة الوصول إليه هو الحصول على المزيد من البيانات. لقد أصبح شعار المديرين الجدد هو: "نثق بالله ــ ونحصل على البيانات من الجميع"، الشعار المتداول في حجيرات وادي السيليكون وفي المصانع وفي طرقات الوكالات الحكومية. إن وجهة النظر في حد ذاتها جيدة، ولكن يمكن أن تضللنا البيانات.

هل يبدو أن التعليم يتدهور؟ قم بوضع اختبارات قياسية لقياس الأداء ومعاقبة المعلمين والمدارس ذات المستويات المتدنية. إن قدرة هذه الاختبارات على قياس

قدرات طلبة المدارس أو جودة التعليم أو احتياجات قوة العمالة الحديثة المبتكرة والقادرة على التكيف أصبحت سؤالاً لا إجابة عنه ـــ ولكنه سؤال لا يمكن للبيانات الإجابة عنه.

هل ترغب في منع الإرهاب؟ قم بإنشاء طبقات من قوائم المراقبة والمنع من السفر جوًّا من أجل السيطرة على الرحلات الجوية. ولكن من غير المؤكد أن تقدم هذه المجموعات البيانية الحماية المتوقعة منها. في واقعة شهيرة، تم منع عضو مجلس الشيوخ الأمريكي عن ولاية ماساتشوسيتس "تيد كينيدي" من السفر جوًّا وخضع للتحقيق لمجرد أن اسمه يماثل اسم شخص موجود في قاعدة البيانات.

لدى الأشخاص العاملين في مجال البيانات تعبير يطلقونه على بعض هذه المشكلات: "القمامة تدخل، القمامة تخرج". في حالات معينة، يكمن السبب في جودة المعلومات الرئيسية، والتي عادة ما يساء استخدام التحليلات الناتجة عنها. في حالة البيانات الضخمة، قد تظهر هذه المشكلات بشكل أكبر أو قد تكون لها عواقب أشد تأثيرًا.

شركة جوجل، كما عرضنا من قبل في الكثير من الأمثلة، تدير كل شيء طبقًا لتحليل البيانات؛ حيث اتضح أن هذه الإستراتيجية هي المسئولة عن الكثير من النجاحات التي حققتها. ولكنها أحيانًا ما تضع الشركة في بعض المشكلات من وقت لآخر. لطالما أصر المؤسسون المشاركون للشركة، "لاري بايج" و"سيرجي برين"، على معرفة درجات التقييم المدرسي للكفاءة الدراسية ومتوسط درجات جميع المتقدمين للعمل في الشركة عند تخرجهم في الجامعة؛ فمن وجهة نظرهم، يقيس الرقم الأول مستوى الكفاءة ويقيس الثاني القدرة على تحقيق النتائج. حتى إن المديرين البارعين في الأربعينيات من عمرهم الذين تقدموا للعمل بالشركة قد طُلب منهم تقديم هذه الأرقام، الأمر الذي أصابهم بالحيرة الشديدة. واصلت الشركة طلب هذه الأرقام فترة طويلة بعد حتى أن أثبتت دراساتها الداخلية أنه لا توجد علاقة بين هذه الأرقام ومستوى الأداء في العمل.

كان على شركة جوجل أن تعرف أكثر؛ حتى لا يغريها سحر البيانات الزائف. كانت هذه القياسات تترك مجالاً محدودًا للتغير في حياة الناس؛ حيث فشل في تقدير المعرفة ونجح في تقدير القدرة على حفظ المواد الدراسية. وقد لا يعكس هذا الأمر مؤهلات المتقدمين للعمل من الناحية الإنسانية؛ حيث يكون قياس البراعة أصعب من قياس العلوم والهندسة. كان هوس شركة جوجل بمثل هذه البيانات من أجل أغراض الموارد البشرية أمرًا غريبًا، خاصة أن مؤسسي الشركة من خريجي مدارس "مونتيسوري" التي تؤكد على التعلم ذاته لا الدرجات. ويكرر هذا الأمر خطأ الشركات التقنية الكبرى عندما اهتمت بالسير الذاتية للمتقدمين لشغل الوظائف أكثر من اهتمامها بقدراتهم الحقيقية. هل كان بمقدور كل من "لاري" و"سيرجي"، الذين لم يكملوا أبحاثهم للحصول على درجة الدكتوراه، أن يصبحوا من مديري الشركة الأسطورية بيل لابس طبقًا لهذه المعايير؟ طبقًا لمعايير شركة جوجل، وليس معايير "بيل جيتس" ولا "مارك زوكربيرج" ولا "ستيف جوبز"، لم يكونوا ليحصلوا على الوظيفة لأنهم لا يحملون درجة أكاديمية.

يبدو أحيانًا أن الشركة تعتمد بشكل مفرط على البيانات؛ حيث قامت "ماريسا ماير"، عندما كانت واحدة من كبار المديرين التنفيذيين بالشركة، بإعطاء أمر إلى فرق العمل لاختبار 41 درجة من اللون الأزرق لتعرف أيًّا من هذه الدرجات يستخدمه الناس بشكل أكبر، وذلك لتحديد لون شريط الأدوات للموقع الإلكتروني الخاص بالشركة. إن رضوخ شركة جوجل للبيانات قد وصل إلى حده الأقصى، حتى إنه قد أشعل شرارة الثورة داخلها.

في عام 2009، استقال أفضل مصممي شركة جوجل "دوجلاس بومان"، في نوبة غضب لأنه لم يعد قادرًا على تحمل القياس الكمي المستمر لكل شيء؛ حيث كتب في إحدى المدونات معلنًا استقالته: "كنت قد تورطت في جدل حول ما إذا كان يجب أن يكون عرض الحدود 3 أو 4 أو 5 بيكسل، وطُلب مني أن أثبت وجهة نظري. لا يمكنني العمل في بيئة عمل مثل هذه. عندما تمتلئ إحدى الشركات بالمهندسين، فإنها تستعين بالمهندسين لحل أية مشكلة تواجهها؛ مما يحصر أي قرار ليصبح مجرد مشكلة منطقية. لقد أصبحت البيانات هي المحرك الرئيسي لجميع القرارات، الأمر الذي سيصيب الشركة بالشلل".

إن العبقرية لا تعتمد على البيانات. ربما تمكن "ستيف جوبز" من التطوير المتواصل لحاسب "ماك" المحمول طوال سنوات بناءً على التقارير الميدانية، ولكنه كان يستخدم حدسه وليس البيانات، لإطلاق أجهزة الآي بود والآي فون والآي باد؛ حيث كان يعتمد على حاسته السادسة؛ حيث قال مقولته الشهيرة، حين أخبر أحد المراسلين بأن أبل لم تقم بأية أبحاث للسوق قبل إطلاق جهاز الآي باد: "ليس من عمل المستهلكين أن يعرفوا ما يريدون".

في كتابه *Seeing Like a State*، يوثق العالم بعلوم الإنسان بجامعة يال "جايمس سكوت"، الطرق التي تتبعها الحكومات، في تبجيلها للقياسات والبيانات، قد انتهى بها الأمر بتحويل حياة الناس إلى جحيم بدلاً من تحسينها؛ حيث كانت تستخدم الخرائط لتحديد كيفية إعادة تنظيم المجتمعات بدلاً من إدراك أي شيء عنهم على أرض الواقع. وكانت تستخدم جداول كبيرة من البيانات عن المحاصيل لتوحيد أسلوب الزراعة دون أن تعرف أي شيء عنها. لقد سلكت جميع الطرق العضوية غير الكاملة التي اتبعها الناس طوال الوقت وربطتها باحتياجاتهم، في بعض الأحيان من أجل إرضاء رغبتها في تحقيق النظام القابل للقياس. إن استخدام البيانات، من وجهة نظر "سكوت"، عادة ما يقوم بزيادة سلطة الأقوياء.

هذه هي ديكتاتورية البيانات في أوضح صورها، والتي تشبه العجرفة التي قادت الولايات المتحدة الأمريكية من جانبها إلى تصعيد الحرب على فيتنام بناءً على عدد قتلى العدو، بدلاً من أن تبني قراراتها على قياسات أكثر منطقية؛ حيث قال " ماكنامارا" في أحد خطاباته عام 1967، حينما كانت الاعتراضات الشعبية على الحرب تتزايد: "لا شك في أنه لا يمكن تبسيط أي موقف إنساني معقد ليصبح مجرد خطوط على الرسم البياني، أو نسب متراصة في جدول، أو أرقام في تقارير الموازنة. ولكن يمكن التفكير بمنطقية في جميع الأمور الواقعية، ولن يكون قياس ما يمكن قياسه بأن تكون راضيًا بأمر ما أقل من المدى الكامل للمنطق". هذا إذا ما

تم استخدام البيانات الصحيحة بالطريقة الصحيحة، وليس من أجل مجرد تمجيد البيانات.

خلال سبعينيات القرن العشرين، قام "روبرت سترينج ماكنامارا" بإدارة البنك العالمي، ثم نَصَّبَ نفسه حمامة السلام في الثمانينيات؛ حيث صرح بانتقاده للأسلحة النووية وأيد حماية البيئة. بعد ذلك، حدث له تحول فكري ونشر مذكراته، تحت عنوان *In Retrospect*، التي انتقدت الفكر الذي أدى للحرب وقراراته التي اتخذها عندما كان وزيرًا للدفاع، حيث كتب فيها: "لقد كنا مخطئين، مخطئين للغاية". ولكنه كان يشير إلى الإستراتيجية العامة للحرب. أما في مسألة البيانات، وبالأخص عدد القتلى، فقد ظل غير نادم على قراراته؛ حيث أقر أن الكثير من إحصاءاته كانت "مضللة أو خاطئة"؛ حيث كتب قائلاً: "ولكن عليك أن تحسب الأمور التي يمكن حسابها، والموت أحد هذه الأمور ...". توفي "ماكنامارا" عام 2009 عن عمر يناهز الثالثة والتسعين عامًا، وكان رجلاً يتمتع بالذكاء لا الحكمة.

قد تغرينا البيانات الضخمة لارتكاب الإثم الذي ارتكبه "ماكنامارا": أن نكون مرتبطين بشكل كبير بالبيانات، وأن نكون مهووسين بالقوة والوعود التي تقدمها، لدرجة أننا نفشل في تقدير حدودها. من أجل أن نكتشف سريعًا مكافئ البيانات الضخمة في حالة أعداد القتلى، كل ما علينا فعله هو تذكر نماذج توقع انتشار الأنفلونزا التي أعدتها جوجل. فكر في موقف، لا يكون مستحيلاً تمامًا، ينتشر فيه نوع مميت من الأنفلونزا في جميع أنحاء البلاد؛ حيث سيشعر الأطباء بالامتنان لوجود قدرة التوقع السريع لأكثر المناطق عرضة للإصابة من خلال قوة طلبات البحث على الإنترنت؛ حيث سيتمكنون من معرفة المناطق التي يجب عليهم التدخل فيها لإعطاء المساعدة.

ولكن افترض أنه في وقت الأزمات؛ حيث يحدث جدل بين القادة السياسيين حول أن معرفة أين من المحتمل أن يتفاقم انتشار المرض وأن محاولة التدخل للمساعدة لا تكفي، فيقومون بفرض حجر صحي عام ــ ليس على جميع سكان تلك المناطق، الأمر الذي لن يكون ضروريًّا ومبالغًا فيه. إن البيانات الضخمة تمكننا من أن نكون أكثر تحديدًا؛ حيث سيسري الحجر الصحي فقط على مستخدمي الإنترنت الذين قاموا بإجراء بحث على الإنترنت يتعلق بإصابتهم بالأنفلونزا. لقد حصلنا ها هنا على البيانات عمن يجب أن نوليهم الاهتمام. ومن ثم يقوم العملاء الفيدراليون المسلحون بقوائم من عناوين بروتوكولات الإنترنت ومعلومات نظام تحديد المواقع العالمي المحمولة، بجمع الأفراد الذين دلت طلباتهم البحثية على إصابتهم بالأنفلونزا ويضعونهم في مركز الحجر الصحي.

ولكن كما قد يبدو هذا السيناريو منطقيًّا لبعض الناس، فإنه خاطئ؛ حيث إن العلاقات التبادلية لا تدل على الأسباب؛ حيث إنه من غير المؤكد أن يكون هؤلاء الأشخاص مصابين بالأنفلونزا؛ لذا يجب أن يخضعوا للفحص؛ حيث قد يصبحون سجناء للتوقعات. ولكن الأمر الأهم هو أنهم قد يقعون ضحية وجهة نظر البيانات التي تفتقد لتقدير المعنى الحقيقي للمعلومات. إن الفكرة الرئيسية من دراسة

نماذج توقع انتشار الأنفلونزا من شركة جوجل هي أن بعض مصطلحات البحث على محركات البحث على الإنترنت لها علاقة تبادلية مع انتشار الوباء ـــ ولكن قد توجد هذه العلاقة التبادلية بسبب بعض الملابسات مثل سماع أحد زملاء العمل يعطس في مكتبه ومن ثم توجه الشخص إلى محرك البحث ليعرف كيف يحمي نفسه من العدوى، ولا يدل هذا على أنه هو نفسه مصاب بالمرض.

الجانب المظلم للبيانات الضخمة

كما رأينا من قبل، تسمح البيانات بوجود المزيد من المراقبة على حياتنا في حين أنها تؤدي إلى جعل بعض الوسائل القانونية في حياتنا والمتعلقة بحماية خصوصيتنا عتيقة الطراز. كما أنها أثبتت عدم فاعلية الأساليب التقنية الرئيسية للحفاظ على هويات المستخدمين مجهلة. الأمر المثير للقلق، هو استخدام توقعات البيانات الضخمة عن الأفراد لعقاب الناس على ميولهم وليس على أفعالهم. الأمر الذي يلغي وجود الإرادة الحرة ويضعف من كرامة الإنسان.

في الوقت ذاته، هناك خوف حقيقي من أن تغري فوائد البيانات الضخمة الناس لأن يطبقوها على التقنيات التي لا تناسبها، أو أن يثقوا بنتائج التحليلات بشكل مبالغ فيه. مع تحسن توقعات البيانات الضخمة، فإن استخدامها يصبح أكثر جاذبية مما يعزز من الهوس بالبيانات حيث إنها قادرة على القيام بالكثير من الأمور. وكانت هذه هي النقمة التي وقعت على "ماكنامارا" والدرس الذي تتضمنه قصته.

علينا أن نحترس من المبالغة في الاعتماد على البيانات بدلاً من تكرار خطأ "إيكاروس"، الذي أعجب كثيرًا بالقدرة التقنية على الطيران التي ابتكرها ولكنه استخدمها بطريقة خاطئة وسقط في البحر. في الفصل التالي، سنستعرض طرقًا يمكننا من خلالها السيطرة على البيانات الضخمة لتفادي سيطرتها علينا.

9
السيطرة

إن تغيير الطرق التي ننتج ونتعامل بها مع المعلومات ستؤدي إلى تغيير القواعد التي نستخدمها في الحكم على أنفسنا، وفي القيم التي يحتاج المجتمع إلى حمايتها. تذكر المثال من فيضان البيانات السابق الذي أفرزه اختراع الطباعة.

قبل اختراع "يوهان جوتنبرج" الطباعة المتحركة في حوالي عام 1450 ميلادية، كان انتشار الأفكار محصورًا لحد كبير على العلاقات الشخصية. وكانت الكتب مقصورة على مكتبات دور العبادة، وكان يحرسها النساك التابعون لدور العبادة بحرص شديد لحمايتها وللحفاظ على الهيمنة عليها. ولكن كان وجود الكتب خارج أسوار دور العبادة نادرًا؛ حيث تمكن عدد قليل من الجامعات من جمع بضع عشرات وربما بضع مئات من الكتب، فقد بدأت جامعة كامبريدج القرن الخامس عشر وهي تمتلك 122 مجلدًا فقط.

بعد بضعة عقود من اختراع "جوتنبرج" آلة الطباعة، تزايد استخدام آلته في ربوع قارة أوربا؛ مما جعل من إنتاج الكتب والمنشورات على نطاق واسع أمرًا ممكنًا. عندما قام "مارتن لوثر" بترجمة النصوص الدينية اللاتينية إلى اللغة الألمانية المتداولة، أصبح الناس يمتلكون سببًا لتعلم القراءة والكتابة: أن يقرءوا هذه الكتب بأنفسهم؛ حيث تمكنوا من عدم اللجوء إلى النساك من أجل تعلم كلمات الله. ومن ثم أصبحت هذه الكتب الدينية الأكثر مبيعًا بين جميع أنواع الكتب. وبمجرد أن تعلموا القراءة والكتابة، واصل الناس القراءة. وقرر بعضهم أن يكتب. في وقت يقل عن متوسط عمر الشخص العادي، تحول تدفق المعلومات من قطرات إلى سيل جارف.

تسبب التغير الجذري في وضع قواعد جديدة للتحكم في الانفجار المعلوماتي الذي أشعلت شرارته الطباعة المتحركة. بعدما وحدت العلمانية قواها، قامت بإنشاء رقابة وتفويض من أجل احتواء الكلمات المطبوعة والسيطرة عليها. وتم ابتكار حقوق الملكية الفكرية من أجل إعطاء الكتاب محفزات قانونية واقتصادية ليبتكروا. بعد ذلك، ضغط المفكرون من أجل وضع قوانين لحماية الكلمات من قمع الحكومات؛ في القرن التاسع عشر، في عدد متزايد من الدول، تحولت حرية التعبير إلى أمر يضمنه دستور البلاد. ولكن أتت هذه الحقوق مصحوبة بالكثير من المسئوليات. بسبب ظهور بعض الصحف الساخرة التي كانت تنتهك خصوصيات الناس أو تشوه سمعتهم، ظهرت قوانين لحماية المجال الشخصي للناس وسمحت لهم بحق رفع دعاوى قضائية بسبب التشهير.

ولكن عكست هذه التغيرات في السيطرة على البيانات تحولاً أكثر عمقًا ورسوخًا في القيم الرئيسية. في ظل اختراع "جوتنبرج" بدأنا في إدراك قوة الكلمة المكتوبة ـــ وفي نهاية الأمر، أهمية المعلومات التي تنتشر في المجتمع بأكمله. بمرور

القرون، اخترنا زيادة كم المعلومات المتدفقة بدلاً من التقليل منه، وأن نحمي أنفسنا بشكل رئيسي من سيطرته علينا من خلال قواعد تحد من إساءة استخدام المعلومات.

مع تحرك العالم في اتجاه استغلال البيانات الضخمة، فإن المجتمع سيخضع لتحول تكويني مشابه. إن البيانات الضخمة تغير بالفعل الكثير من سمات حياتنا وطرق تفكيرنا؛ مما يجبرنا على إعادة التفكير في المبادئ الرئيسية لكيفية تشجيع نموها والتخفيف من قدرتها على إلحاق الأذى. ولكن على العكس من أسلافنا إبان فترة ثورة الطباعة وما قبلها؛ حيث إننا لن تكون أمامنا قرون طويلة لنقوم فيها بالتعديل من أوضاعنا، بل عدة سنوات.

إن التغيرات البسيطة التي ستدخل على القواعد الحالية لن تكون كافية للسيطرة على عصر البيانات الضخمة وعلى مزاجية الجانب المظلم لها. بدلاً من التغيير القياسي للقواعد، يدعونا الوضع إلى إحداث تغيير فكري. إن حماية الخصوصية تتطلب مستخدمي بيانات ضخمة أكثر تحملاً لمسئولية أفعالهم. وفي الوقت ذاته، سيعيد المجتمع تعريف فكرة العدالة لضمان حرية الأفراد في القيام بأفعالهم (وبالتالي يتحملون مسئولية أفعالهم). وفي النهاية، يجب أن يظهر نوع جديد من المؤسسات والمتخصصين الذين سيتمكنون من تفسير الخوارزميات المعقدة التي تكمن خلف نتائج البيانات الضخمة، ومن أجل أن يقدموا النصح للأشخاص الذين قد تسبب لهم البيانات الأذى.

من الخصوصية إلى تحمل المسئولية

لعقود طويلة، سمح المبدأ الرئيسي لقوانين الخصوصية في جميع أنحاء العالم للأفراد بأن يكونوا في موضع التحكم من خلال السماح لهم بأن يقرروا ما إذا كان ستتم معالجة معلوماتهم الشخصية وكيف سيتم ذلك ومن سيقوم بذلك. في عصر الإنترنت، تحول هذا المبدأ ليصبح نظامًا منهجيًّا من "الملاحظة والقبول". في عصر البيانات الضخمة؛ حيث تكمن أغلب قيمة البيانات في استخداماتها الثانوية التي لم توضع في الاعتبار عند جمع هذه البيانات في المقام الأول، فإن مثل هذه الآليات للحفاظ على الخصوصية لن تكون ملائمة.

إننا نتخيل إطار عمل يختص بحماية الخصوصية يختلف تمامًا عن عصر البيانات الضخمة، إطار عمل يركز بشكل أقل على قبول المستخدم لجمع بياناته ويركز أكثر على تحميل مستخدمي البيانات مسئولية ما يفعلونه بها. في مثل هذا العالم، عادة ما ستقوم الشركات بتقييم إعادة استخدام البيانات بناءً على تأثيرها على الأفراد الذين تتم معالجة معلوماتهم الشخصية. يجب ألا يتم هذا الأمر بأدق التفاصيل في جميع الحالات؛ حيث إن قوانين حماية الخصوصية في المستقبل ستحدد فئات أوسع من الاستخدامات، من بينها الاستخدامات التي يمكن السماح بها دون إجراءات وقائية أو مع وجود إجراءات وقائية محدودة ومعيارية. في واحدة من المبادرات المغامرة، قد يضع المشرعون قوانين لكيفية تقييم مستخدمي البيانات لمخاطر أحد استخداماتها وتحديد ما يمكن فعله لتجنب الأضرار أو على الأقل التقليل منها.

يشجع هذا الأمر على ابتكار استخدامات جديدة للبيانات، ولكن في الوقت ذاته، يضمن وجود معايير كافية يتم اتباعها للتأكد من عدم الإضرار بأي أحد.

إن إجراء تقييم استخدامات البيانات الضخمة الرسمي بشكل صحيح وتطبيق نتائجه بدقة يقدم فوائد ملموسة لمستخدمي البيانات: سيكونون أحرارًا في سعيهم وراء التوصل إلى الاستخدامات الثانية للبيانات الشخصية في الكثير من الحالات دون الاضطرار للعودة إلى الأفراد للحصول على موافقتهم. ومن ناحية أخرى، فإن التقييم الرديء أو التطبيق السيئ لإجراءات الوقاية قد يعرض مستخدمي البيانات إلى المساءلة القانونية، والعقوبات التشريعية مثل فرض الوصاية أو الغرامات وربما حتى التعرض للمحاكمات الجنائية. إن مسئولية مستخدمي البيانات ستصلح فقط عندما تكتمل.

لكي نرى كيف يمكن أن يتحقق هذا الأمر على أرض الواقع، استرجع مثال تحويل معلومات المؤخرات إلى بيانات من الفصل الخامس. تخيل لو أن إحدى الشركات قد باعت نظامًا لحماية السيارات من السرقة والذي يستخدم معلومات طريقة جلوس السائق كمحدد لشخصية من يقود السيارة. بعد ذلك، يقوم النظام بإعادة تحليل المعلومات من أجل توقع "مدى انتباه" قائد السيارة، مثل أن يكون مرهقًا أو مخمورًا أو غاضبًا؛ حيث يقوم بإرسال إشارة إلى قائدي السيارات الأخرى لينتبهوا لتفادي وقوع حوادث. بموجب قوانين حماية الخصوصية الحالية، قد تعتقد الشركات أنها بحاجة إلى الحصول مجددًا على ملاحظة المستخدمين وموافقتهم لأنها لم تحصل منذ البداية على الموافقة باستخدام البيانات بهذه الطريقة. ولكن بموجب نظام تحمل مستخدم البيانات للمسئولية حيال استخدامه إياها، قد تقوم الشركة بتقييم مخاطر الاستخدام المحتمل، وإن وجدت تلك المخاطر في أدنى مستوياتها فلا بأس من الاستمرار في تنفيذ خططها ـــ وتحسين مستوى أمان الطرق خلال هذه العملية.

إن تحويل تحمل المسئولية من على عاتق العامة إلى عاتق مستخدمي البيانات لن يكون مطلبًا قانونيًّا لمحو المعلومات الشخصية بمجرد أن ينتهي الغرض الأولي لجمعها، كما تتطلب أغلب قوانين حماية الخصوصية الحالية. إن هذا التحول يعتبر من التحولات المهمة؛ حيث إنه من خلال استغلال القيمة الكامنة للبيانات فقط يمكن لأصحاب العقول المبتكرة أمثال "موراي " أن يزدهروا من خلال اعتصار البيانات لاستخراج أكبر قيمة ممكنة منها لصالحهم ـــ ولصالح المجتمع. ولكن، سيكون من المسموح لمستخدمي البيانات بأن يحتفظوا بالمعلومات الشخصية وقتًا أطول، ولكن ليس للأبد؛ حيث إن المجتمع بحاجة إلى المقارنة الدقيقة لفوائد إعادة استخدام البيانات مقابل مخاطر نشرها الزائد عن الحد.

من أجل الوصول إلى التوازن الصحيح، قد يختار واضعو النظم أطرًا زمنية مختلفة لإعادة الاستخدام بناءً على المخاطر الملازمة للبيانات، وكذلك بناءً على قيم المجتمعات المختلفة. قد تتعامل بعض الدول مع الأمر بحذر أكبر من دول أخرى؛ حيث إن بعض أنواع البيانات قد تكون حساسة أكثر من بعضها الآخر. من شأن هذا الأسلوب أن يبعد وهم "الذاكرة الدائمة" ـــ الخطر القائم بأن المرء لن يتمكن أبدًا

من الهرب من ماضيه حيث إن السجلات الرقمية دائمًا ما ستذكره بذلك، وإلا فستظل بياناتنا الشخصية تثقل كاهلنا كما لو كانت "سيف داموكليس"، تهددنا بأن تواصل طعننا لسنوات بفعل بعض التفاصيل عن هوياتنا أو قيامنا بشراء شيء ندمنا عليه. تحفز الحدود الزمنية أيضًا مالكي البيانات على استخدامها قبل أن يفقدوها، وهذا ما يحقق ما نعتقد أنه التوازن المثالي لعصر البيانات الضخمة: تحصل الشركات على حق استخدام البيانات الشخصية لفترة أطول، ولكن عليها في المقابل أن تتحمل مسئولية استخدامها إياها، بالإضافة إلى الالتزام بمحو تلك البيانات بعد مرور فترة محددة من الوقت.

إلى جانب التحول التنظيمي من "حماية الخصوصية من خلال الموافقة" إلى "حماية الخصوصية من خلال تحمل المسئولية"، يمكننا تخيل بعض الابتكارات التقنية للمساعدة على حماية الخصوصية في حالات بعينها. هناك أسلوب جديد وهو مفهوم "الخصوصية المتفاوتة": وهو تعمد التعتيم على البيانات بحيث لا يكشف البحث داخل المجموعات البيانية الكبيرة عن النتائج الدقيقة، بل يكشف عن نتائج تقريبية؛ مما يجعل من الصعب الربط بين نقاط بيانية معينة وهويات المستخدمين، بل وزيادة تكلفة هذا الأمر أيضًا.

يبدو التعتيم على المعلومات كما لو كان سيدمر الأفكار القيمة. ولكن لا يجب بالضرورة أن يحدث هذا الأمر ــ أو على أقل تقدير، قد تكون الاحتمالات مبشرة. على سبيل المثال، لاحظ خبراء السياسات التقنية أن شركة فيس بوك تعتمد على نوع من الخصوصية المتفاوتة عندما تقدم تقارير عن مستخدميها إلى المعلنين المحتملين: حيث تكون الأرقام التي تقدمها للمعلنين تقريبية؛ لذا لا يمكن من خلالها تحديد هويات المستخدمين، فعند البحث عن عبارة السيدات الآسيويات من أتلانتا اللواتي يهتممن بيوجا "أشاتانجا"، ستظهر النتائج "حوالي 400" وليس الرقم الواقعي؛ مما يجعل من المستحيل استخدام المعلومات الناتجة لتضييق مجال البحث إحصائيًّا لتحديده على شخص واحد فقط.

إن التحول من موافقة الأفراد إلى تحمل مستخدمي البيانات لمسئوليتها يعتبر تحولاً أساسيًّا وضروريًّا للسيطرة الفعالة على البيانات، ولكنه ليس الأسلوب الوحيد.

الناس مقابل التوقعات

تحمل المحاكم الناس مسئولية أفعالهم. عندما يعلن القضاة أحكامهم غير المنحازة بعد محاكمة عادلة، فقد تم تحقيق العدالة. ولكن في عصر البيانات الضخمة، نحتاج إلى إعادة تعريف فكرتنا عن العدالة من أجل الحفاظ على فكرة حرية البشر: الإرادة الحرة التي من خلالها يختار البشر أفعالهم. إنها الفكرة البسيطة التي تقول إنه يمكن للأفراد ويجب عليهم أن يتحملوا مسئولية سلوكياتهم، وليس عن ميولهم.

قبل البيانات الضخمة، كانت هذه الحرية التي لا يمكن الاستغناء عنها واضحة لدرجة أنه لم تكن هناك حاجة للتعبير عنها؛ حيث إن هذه هي الطريقة التي يعمل بها نظامنا القانوني: تحميل الناس مسئولية أفعالهم بعد تقييم ما قاموا به. ولكن على

النقيض، مع وجود البيانات الضخمة يمكننا آن نتوقع أفعال البشر بدقة متزايدة؛ مما يغرينا بمحاسبة الناس ليس على ما فعلوه ولكن على ما توقعنا أن يقوموا به.
في عصر البيانات الضخمة، سيكون لزامًا علينا أن نوسع إدراكنا لمفهوم العدالة، وتتطلب أن تتضمن إجراءات وقائية لحماية الحرية البشرية بنفس درجة حمايتنا الحالية للعدالة الإجرائية. بدون هذه الإجراءات الوقائية ستضعف فكرة العدالة بدرجة كبيرة.
من خلال ضمان الحرية البشرية، سنضمن أن تقرر الحكومات أحكامها على سلوكياتنا بناءً على أفعالنا الحقيقية، وليس على تحليل البيانات الضخمة. وبالتالي، سيكون على الحكومات أن تحملنا مسئولية أفعالنا الماضية، وليس مسئولية التوقعات التحليلية لأفعالنا المستقبلية. وعندما تحكم الدولة على أفعالنا السابقة، فإنها ستمتنع عن الاعتماد على البيانات الضخمة وحدها. على سبيل المثال، تأمل قضية الاحتكار التي تورطت فيها تسع شركات. من المقبول تمامًا أن يتم استخدام البيانات الضخمة لتحديد المتواطئين المحتملين بحيث يتمكن واضعو النظم من إجراء التحقيقات ورفع دعوى قضائية بالطرق التقليدية. ولكن لا يمكن عقاب هذه الشركات فقط لأن البيانات الضخمة اقترحت أنها من المحتمل أن تكون قد ارتكبت جرمًا.
هناك مبدأ آخر يجب تطبيقه خارج الأطر الحكومية، عندما تتخذ الشركات قرارات غاية في الأهمية عنا ـــ التوظيف أو الاستغناء أو تقديم عرض بالرهن أو إلغاء إحدى البطاقات الائتمانية. إذا ما بُنيت هذه القرارات بشكل أكبر على توقعات البيانات الضخمة، فإننا ننصح بوجوب وجود بعض الإجراءات الوقائية. أولاً المصارحة: جعل البيانات والمعادلات الخوارزمية التي بُنيت عليها التوقعات التي تؤثر على شخص ما متوافرة. ثانيًا التوثيق: وجود المعادلات الخوارزمية المصدق على صلاحيتها للاستخدامات الحساسة من خبير محايد مشهود له بالصلاحية والنزاهة. القدرة على النقض: تحديد طرق ثابتة يمكن من خلالها للناس أن يدحضوا تلك التوقعات المتعلقة بهم. (يشبه هذا الأمر المبدأ الساري في مجال العلوم الذي ينص على الإفصاح عن أية عوامل قد تؤثر بالسلب على نتائج إحدى الدراسات) .
الأمر الأهم، إن ضمان الحرية البشرية يحمينا من تهديد ديكتاتورية البيانات، التي نمنح من خلالها البيانات معنى وأهمية أكثر مما تستحق.
من المهم أيضًا أن نحمي المسئولية الفردية؛ حيث قد يواجه المجتمع إغراءً شديدًا للتوقف عن تحميل الأفراد المسئولية عن أفعالهم ويتحول بدلاً من ذلك إلى السيطرة على المخاطر، أي أن يبني قراراته حول الناس على أساس تقييم احتمالات وأرجحيات النتائج المحتملة. مع توافر الكثير من البيانات الموضوعية؛ فقد يبدو أنه من الجيد أن نخلي عملية اتخاذ القرار من العاطفية والفردية، وأن نعتمد على الخوارزميات بدلاً من تقييم الأفعال بواسطة القضاة والمقيّمين، وأن نضع القرارات في إطار لا يتعلق بالمسئولية الشخصية ولكن يتعلق أكثر بالمخاطر "الموضوعية" وكيفية تفاديها.
على سبيل المثال، تقدم البيانات الضخمة دعوة قوية لتوقع أي من الأشخاص من

المحتمل أن يرتكبوا الجرائم وإخضاعهم لمعاملة خاصة، وأن يوضعوا تحت المراقبة مرارًا وتكرارًا تحت شعار التقليل من المخاطر. قد يشعر الأشخاص الذين يتم تصنيفهم بهذه الطريقة بأنهم يعاقبون دون أن تتم مواجهتهم بسلوكياتهم الفعلية ودون أن يتحملوا مسئوليتها. تخيل تحديد واحدة من المعادلات الخوارزمية لأحد المراهقين على أنه من المحتمل جدًّا أن يرتكب مخالفة في خلال السنوات الثلاث القادمة. ونتيجة لهذا، تعين السلطات أحد موظفي الخدمات الاجتماعية لزيارته مرة كل شهر، من أجل مراقبته ومساعدته على الابتعاد عن المشاكل.

إذا ما كان المراهق أو أقاربه أو أصدقاؤه أو معلموه أو من يعمل لديهم رأوا هذه الزيارات على أنها أمر مشين، كما سيرونها بكل تأكيد، فإن هذا التدخل سيحمل سمات العقاب، العقاب على فعل لم يحدث. ولن يختلف الوضع كثيرًا إذا لم تعتبر هذه الزيارات عقابًا بل مجرد محاولة للتقليل من احتمالات حدوث مشكلات في المستقبل ــ كطريقة للتقليل من المخاطر (في هذه الحالة، خطر الجريمة الذي من شأنه أن يؤثر بالسلب على الأمن العام). كلما زاد التحول من تحميل الناس مسئولية أفعالهم إلى الاعتماد على التدخلات المبنية على البيانات للتقليل من المخاطر التي يتعرض لها المجتمع، قل تقديرنا لمبدأ المسئولية الفردية. إن الدولة التوقعية هي الدولة الحاكمة، وبعضها بالفعل على هذا المنوال. إن حرمان الناس من مسئوليتهم عن أفعالهم يدمر حريتهم الأساسية لاختيار سلوكهم.

إذا ما حددت الدولة اتخاذها للكثير من القرارات بناءً على التوقعات والرغبة في التخفيف من المخاطر، فإن خياراتنا الفردية ــ وبالتالي حريتنا للتصرف ــ لن تكون مهمة. بدون الشعور بالذنب لن تكون هناك براءة. إن الاستسلام لهذا الأمر من شأنه ألا يحسن من مجتمعنا، بل من شأنه أن يفقره أكثر.

إن الركيزة الرئيسية لحكم البيانات الضخمة يجب أن تكون ضمانة لأننا سنستمر في الحكم على الناس من خلال الوضع في الاعتبار مسئولياتهم الشخصية وسلوكياتهم الحقيقية، ليس من خلال معالجة البيانات "بموضوعية" من أجل تحديد ما إذا كان من المحتمل أن يقوموا بأفعال خاطئة. بهذه الطريقة وحدها يمكننا أن نعاملهم كبشر: كأشخاص يمتلكون الحرية لاختيار أفعالهم والحق في أن يحكم عليهم من خلالها.

كسر الصندوق الأسود

تقيم أنظمة الحاسب في الوقت الحالي قراراتها بناءً على القواعد التي تمت برمجتها على اتباعها؛ لذا، عندما يثبت أن أحد القرارات خطأ، كما من المحتم أن يحدث من وقت لآخر، يمكننا أن نعود ونرى ما الذي جعل الحاسب يخطئ. على سبيل المثال، يمكننا أن نحقق في المسائل على غرار: "لماذا قام نظام الطيار الآلي بزيادة ميل الطائرة بمقدار خمس درجات لأعلى عندما قرأ الحساس الخارجي وجود زيادة مفاجئة في الرطوبة؟". يمكننا في الوقت الحالي أن نفتح شفرات الحواسب ونقوم بفحصها، ويمكن للأشخاص القادرين على ترجمتها أن يتتبعوا ويستوعبوا الأساسيات التي تقوم عليها قراراتها، مهما زادت درجة تعقيدها.

مع وجود تحليل البيانات الضخمة، ستصبح هذه القدرة على التتبع أصعب كثيرًا؛ حيث إن الأساس الذي تقوم عليه توقعات الخوارزميات قد تكون معقدة بدرجة كبيرة لا تسمح لأغلب البشر بأن يستوعبوها.

عندما كانت الحواسب تبرمج لتتبع مجموعات التعليمات، كما حدث مع برنامج آي بي إم القديم لترجمة اللغة الروسية إلى الإنجليزية عام 1954، تمكن البشر من فهم الكيفية التي يحول بها البرنامج الكلمات إلى كلمات أخرى. ولكن يستخدم برنامج الترجمة من شركة جوجل مليارات الصفحات المترجمة للخروج بحكمه على ما إذا كانت كلمة "Light" الإنجليزية تعادل "ضوء" أو "خفيف" بالعربية (أي ما إذا كانت الكلمة تشير إلى السطوع أم الوزن). من المستحيل أن يتمكن البشر من تتبع الأسباب الدقيقة لاختيار البرنامج للكلمات لأنها تعتمد على كميات هائلة من البيانات والحسابات الإحصائية الضخمة.

إن البيانات الضخمة تعمل على مقياس يفوق استيعابنا العادي. على سبيل المثال، العلاقة التبادلية التي اكتشفتها جوجل ما بين كمية قليلة من المصطلحات البحثية وانتشار وباء الأنفلونزا كانت نتاج اختبار 450 مليونًا من النماذج الرياضية. على النقيض، قامت "سينثيا رودين" بتصميم 106 نماذج لتوقع ما إذا كانت فتحات الصرف قد تشتعل فيها النيران أم لا، كما تمكنت من أن تشرح لمديري شركة "كون إديسون" سبب تحديد برنامجها لأولوية المواقع التي تجب صيانتها كما فعل. "إمكانية التفسير"، كما يُطلق عليها في دوائر الذكاء الاصطناعي، مهمة بالنسبة لنا نحن البشر الذين نميل إلى معرفة السبب وليس الشيء وحده. ولكن ماذا لو تمكن النظام من إنتاج 601 من النماذج للتوقع بدلاً من مجرد 106، فأي من تلك الأولويات الهائلة كانت أهميتها ضئيلة, ولكنها ــ عند دمجها معًا ــ عملت على تحسين دقة النموذج بأكمله؟ إن الأساس الذي تقوم عليه أي من التوقعات يمكن أن يكون معقدًا بدرجة مذهلة. ما الذي أخبرت به المديرين لتقنعهم بأن يعيدوا توزيع ميزانياتهم المحدودة؟

في هذه السيناريوهات، يمكننا أن نرى تحول المخاطر التي تنتجها توقعات البيانات الضخمة، والخوارزميات والمجموعات البيانية التي تنتجها، إلى صناديق سوداء لا تقدم لنا أية محاسبة أو قدرة على تتبعها أو ثقة. لمنع هذا الأمر، ستتطلب البيانات الضخمة متابعة وشفافية، والتي ستتطلب بدورها أنواعًا جديدة من الخبرات والمؤسسات. سيقدم هؤلاء الوافدون الجدد الدعم في مجالات يحتاج فيها المجتمع إلى فحص توقعات البيانات الضخمة وتمكين الناس الذين يشعرون بأنهم مخطئون من تصحيح أخطائهم.

إننا كمجتمع، لطالما رأينا مثل هذه الكيانات الجديدة تظهر عند حدوث زيادة مفرطة في تعقيد وتخصص مجال بعينه مما يخلق حاجة ملحة لوجود خبراء قادرين على ترويض التقنيات الجديدة. لقد مرت مهن مثل القانون والطب والحسابات والهندسة بهذا التحول منذ أكثر من قرن مضى. وفي الوقت الحاضر، ظهر المتخصصون في أمن الحواسب للتأكيد على أن الشركات تنصاع للممارسات الفضلى التي حددتها كيانات على غرار المؤسسة الدولية للمعايير (والتي تكونت

من أجل تلبية حاجات جديدة لوجود إرشادات في هذا المجال).

ستتطلب البيانات الضخمة وجود مجموعة جديدة من الأشخاص ليلعبوا هذا الدور. ربما سيُطلق عليهم "الخوارزميون". وقد يظهرون في صورتين ـــ كيانات مستقلة ليتابعوا الشركات من الخارج، وموظفين أو إدارات ليتابعوا الشركات من الداخل ـــ مثلما تمتلك الشركات محاسبين يعملون فيها وكذلك مراجعين من الخارج لمراجعة حساباتها.

ظهور متخصصي الخوارزميات

قد يشغل هذه المهن خبراء من مجالات علوم الحاسب والرياضيات والإحصاء؛ وقد يعملون كمراقبين لتحليلات وتوقعات البيانات الضخمة. قد يتعهد متخصصو الخوارزميات بعدم التحيز والحفاظ على سرية البيانات، كما يفعل المحاسبون وبعض المتخصصين في الوقت الحالي. قد يقومون بتقييم مصادر البيانات، واختيار أدوات التحليل والتوقع بما فيها الخوارزميات والنماذج، وكذلك تفسير النتائج. في حالة وجود نزاع، يمكنهم الوصول إلى الخوارزميات والمقاربات الإحصائية والمجموعات البيانية التي قام عليها اتخاذ قرار بعينه.

إن كان يوجد متخصص في الخوارزميات ضمن فريق عمل إدارة الأمن القومي عام 2004، لكان منع الوكالة من إصدار قائمة منع من السفر مليئة بالأخطاء لدرجة منع عضو مجلس الشيوخ الأمريكي السيناتور "كينيدي" من السفر. من بين الحالات الحديثة التي لعبت فيها الخوارزميات الضخمة دورًا مهمًّا، ما حدث في كل من اليابان وفرنسا وألمانيا وإيطاليا، والتي اشتكى فيها الناس من أن سمة "الاستكمال التلقائي" من شركة جوجل، والتي تعمل على إنشاء قائمة بمصطلحات البحث الشائعة التي تتعلق بالكلمات المكتوبة في محرك البحث، قد شهرت بهم. كانت هذه القوائم تعتمد بشكل كبير على تكرار الأبحاث السابقة عن المصطلحات نفسها: يتم ترتيب المصطلحات وفقًا لاحتمالاتها الرياضية. ولكن، من منا لن يغضب إذا ما ظهرت كلمة "مجرم" بجانب اسمه عندما تبحث أنماط البحث العملية أو الرومانسية في شبكة الإنترنت لتحصل على نتائج متعلقة بنا؟

إننا نتخيل الخوارزميات على أنها مقاربات لحل المشكلات التي تتعلق بالسوق مثل تلك التي تغير من التشريعات التي تتدخل بشكل كبير في شئون البشر. إنها ستلبي الحاجة نفسها التي ملأها المحاسبون والمراجعون عندما ظهروا في أوائل القرن العشرين للتعامل مع الكم الكبير من المعلومات المالية. لقد كان الهجوم العددي صعبًا على إدراك الناس؛ لدرجة أن تطلب وجود متخصصين منظمين بطرق ذكية وذاتية التنظيم. وكان رد فعل السوق على هذا الأمر هو زيادة أهمية القطاع الجديد من الشركات المتنافسة في مجال الرقابة المالية. من خلال تقديم هذه الخدمة، دعم هذا النوع الجديد من المتخصصين ثقة المجتمع بعلماء الاقتصاد. قد تستفيد البيانات الضخمة من الثقة المماثلة التي قد يقدمها متخصصو الخوارزميات، بل ويجب ذلك.

متخصصو الخوارزميات الخارجيون

إننا نتخيل أن يكون متخصصو الخوارزميات الخارجيون مراجعين غير منحازين لمراجعة دقة وصلاحية توقعات البيانات الضخمة عند حاجة الحكومة إلى هذه المراجعات، مثل أن تكون بموجب أمر قضائي أو تشريعي. كما أنهم قد يتخذون من شركات البيانات الضخمة عملاء؛ حيث يقومون بعمل مراجعات لصالح الشركات التي تحتاج إلى الحصول على دعم من متخصصين. كما قد يقومون بالتأكيد على كفاءة تطبيقات البيانات الضخمة مثل التقنيات المضادة للتزوير أو أنظمة تجارة السندات. وفي النهاية، يكون متخصصو الخوارزميات الخارجيون على استعداد دائم للتشاور مع الوكالات الحكومية على الاستخدامات المثلى للبيانات الضخمة في القطاع العام.

كما حدث في الطب والقانون وغيرها من المهن، يمكننا أن نتخيل أن تنظم هذه المهنة نفسها من خلال وضع قانون للعمل. وسيتم تنفيذ معايير حيادية وسرية وكفاءة واحترافية متخصصي الخوارزميات طبقًا لقواعد صارمة، وإذا ما فشلوا في تطبيق هذه المعايير، فسيكونون عرضة للدعاوى القضائية. يمكن أيضًا أن يتم استدعاؤهم كشهود متخصصين في المحاكمات، أو يعملون على أنهم "سادة المحاكمات"، الخبراء الذين يعينهم القضاة لمساعدتهم في الأمور التقنية في القضايا المعقدة.

علاوة على ذلك، يمكن للأشخاص الذين يعتقدون أن البيانات الضخمة قد سببت لهم الأذى ــ المريض الذي تم رفض خضوعه لعملية جراحية، السجين الذي حرم من إطلاق السراح المشروط، طالب القرض الذي رُفض رهن عقاره ــ أن يفحصوا الخوارزميات مثلما فعلوا مع المحامين لمساعدتهم على فهم هذه القرارات ونقضها.

متخصصو الخوارزميات الداخليون

يعمل متخصصو الخوارزميات الداخليون داخل المؤسسات لمتابعة أنشطة البيانات الضخمة داخلها؛ حيث لا يراقبون ما يهم الشركة فقط، بل يراقبون أيضًا اهتمامات الأشخاص الذين تؤثر فيهم تحليلات البيانات الضخمة. ويقومون بمتابعة عمليات البيانات الضخمة ويعتبرون حلقة الوصل الأولى مع أي شخص يشعر بأن توقعات مؤسسة البيانات الضخمة قد سببت له الضرر. كما أنهم يدققون في تحليلات البيانات الضخمة من أجل تحري النزاهة والدقة قبل العمل بها. للقيام بأولى هاتين المهمتين، يجب أن يمتلك متخصصو الخوارزميات مستوى معينًا من الحرية والحيادية داخل المؤسسة التي يعملون بها.

قد تبدو فكرة أن يظل الموظف الذي يعمل لحساب شركة ما محافظًا على حياديته فيما يتعلق بعملياتها أمرًا منافيًا للمنطق، ولكن مثل هذه الحالات تعتبر شائعة بشكل جيد. من بين الأمثلة على هذا الأمر، أقسام المراقبة في المؤسسات المالية الكبرى، وكذلك مجلس إدارة الكثير من الشركات، والمسئولون عن التعامل مع حاملي الأسهم ولا تقع إدارة الشركة ضمن نطاق مسئولياتهم. والكثير من الشركات الإعلامية، من بينها صحف نيويورك تايمز وواشنطن بوست، التي

توظف محققين في شكاوى العملاء الذين تقوم مهمتهم الأساسية على صون ثقة العامة بالشركة. يقوم هؤلاء الموظفون بالتعامل مع شكاوى القراء وعادة ما يعاقبون أصحاب العمل على الملأ عندما يكتشفون أنهم قد أخطأوا.

هناك تشبيه أقرب من هذا لمتخصصي الخوارزميات الداخليين ـــ متخصص مسئول عن التأكد من أن المعلومات الشخصية لا يُساء استخدامها من قبل المؤسسة. على سبيل المثال، تتطلب ألمانيا مؤسسات تتخطى حجمًا معينًا (بوجه عام عشرة أشخاص على الأقل يعملون على معالجة المعلومات الشخصية) لتعيين ممثل لحماية البيانات. منذ سبعينيات القرن العشرين، طور هؤلاء الممثلون داخل الشركات العقيدة الأخلاقية العملية وروح التضامن بين الموظفين؛ حيث يجتمعون بشكل منتظم من أجل مناقشة الممارسات المثلى والتدرب والحصول على وسائل الإعلام المتخصصة ومؤتمراتهم الخاصة. علاوة على هذا، نجحوا في الحفاظ على تعهداتهم لموظفيهم ولواجباتهم كمراجعين محايدين، ونجحوا في العمل كمحققين للحفاظ على البيانات في أثناء ترسيخهم قواعد خصوصية المعلومات من خلال أعمال شركاتهم. نعتقد أن الخوارزميات التي تعمل داخل الشركات يمكنها القيام بالمثل.

السيطرة على سادة البيانات

إن أهمية البيانات بالنسبة لمجتمع المعلومات تماثل أهمية الوقود للاقتصاد الصناعي: المصدر الأهم الذي يغذي المبتكرات التي يعتمد عليها الناس. بدون وجود مصدر غني ونشط للبيانات وسوق قوية للخدمات، فقد تخمد الابتكارات والإنتاجية المحتملة.

في هذا الفصل، وضعنا ثلاث إستراتيجيات رئيسية جديدة للسيطرة على البيانات الضخمة تتعلق بالخصوصية والتوجهات ومرجعية الخوارزميات. إننا نثق بأنه مع وجود هذه الإستراتيجيات موضع التطبيق، سنتمكن من الحد من الجانب المظلم للبيانات الضخمة. ولكن مع تطور مجال البيانات الضخمة الوليد، سيظهر المزيد من التحديات المهمة التي يجب وضع إجراءات احترازية لها من أجل المحافظة على أسواق البيانات الضخمة التنافسية. علينا أن نمنع ظهور سادة للبيانات في القرن الحادي والعشرين، المكافئ الحديث لسادة الاحتكار من القرن التاسع عشر الذين هيمنوا على السكك الحديدية وصناعة الصلب وشبكات التلغراف في أمريكا.

من أجل السيطرة على رجال الصناعة الأوائل هؤلاء، وضعت الولايات المتحدة الأمريكية قوانين مضادة للاحتكار قابلة للتكيف مع تغير الأوضاع بشكل كبير. القوانين التي وضعت في الأصل من أجل خطوط السكك الحديدية في أوائل القرن التاسع عشر، تم تطبيقها فيما بعد على الشركات التي كانت تعتبر الحارس الأمين لتدفق المعلومات التي اعتمد عليها عالم الأعمال، من سجل الخزينة القومي في العقد الأول من القرن العشرين، إلى شركة آي بي إم في الستينيات وما يليها، وزيروكس في السبعينيات، وإيه تي آند تي في الثمانينيات، ومايكروسوفت في التسعينيات، وجوجل في الوقت الحالي. لقد أصبحت التقنيات التي ابتكرتها هذه

الشركات المكون الرئيسي "للبنية التحتية المعلوماتية" للاقتصاد، وتطلبت وجود قوة القانون لمنع احتكارها.

من أجل توفير الظروف الملائمة لوجود سوق حية للبيانات الضخمة، فإننا نحتاج إلى وجود معايير مشابهة لتلك التي أسست مبادئ المنافسة والرقابة على تلك المجالات التقنية الأولى. يجب علينا أن نؤكد على التعاملات البيانية، مثلاً من خلال حقوق الاستغلال أو القابلية للتشغيل البيني. يعمل هذا الأمر على إبراز مشكلة ما إذا كان المجتمع قد يستفيد من "حق استثناء" البيانات المصاغ جيدًا والمتوازن (الذي يشبه حق الملكية الفكرية، وأن يكون على القدر نفسه من التحدي الذي يبدو عليه). قد يكون تحقيق هذا الأمر مهمة صعبة على صناع السياسات ــ ومليئًا بالمخاطر بالنسبة لبقيتنا.

من المستحيل أن نتنبأ بكيفية تطور التكنولوجيا، حتى إن البيانات الضخمة ذاتها لن تكون قادرة على توقع كيفية تطورها. وعلى المنظمين لهذا الأمر أن يوازنوا بين التصرف بحذر وجرأة ــ ويشير تاريخ قانون منع الاحتكار إلى وجود طريقة واحدة لتحقيق هذا الأمر.

لقد كبح قانون منع الاحتكار القوى المسيئة. ولكن المثير للدهشة، أن مبادئه قد انتقلت بسلاسة من قطاع لآخر، ومرت بمختلف أنواع صناعات الشبكات. هذا هو نوع التشريعات القوية ــ التي لا تحابي أحد أنواع التكنولوجيا على غيره من التكنولوجيات الأخرى ــ المفيدة؛ حيث إنها تعمل على حماية المنافسة دون الادعاء بقيامها بأي أمر آخر؛ لذا فإن قوانين منع الاحتكار من شأنها أن تعطي دفعة قوية للبيانات الضخمة مثلما فعلت مع خطوط السكك الحديدية من قبل. هذا بالإضافة إلى أن بعض أكبر مالكي البيانات في العالم وكذلك الحكومات، سيكون عليها أن تنشر بعضًا من البيانات التي تملكها على الملأ. الأمر المشجع هو أن بعضها بدأ بالفعل في القيام بهذين الأمرين ــ على الأقل لحد ما.

الدرس المستفاد من تشريعات منع الاحتكار هو أنه بمجرد تحديد المبادئ العامة، يقوم المشرعون بتنفيذها للتأكيد على المستوى الصحيح من الإجراءات الاحترازية والدعم. وتقوم الإستراتيجيات الثلاث التي وضعناها سابقًا بالمثل ــ التحول بحماية الخصوصية من السماح الفردي إلى مسئولية مستخدمي البيانات، الحفاظ على الحرية البشرية من التوقعات، اختراع طائفة جديدة من مراجعي البيانات الضخمة الذين نطلق عليهم "متخصصو الخوارزميات" ــ بأن تعمل كحجر الأساس للسيطرة الفعالة والعادلة على المعلومات في عصر البيانات الضخمة.

في الكثير من المجالات، من التكنولوجيا النووية إلى الهندسة الإحيائية، نقوم في البداية ببناء الأدوات التي نكتشف أنها قد تسبب الضرر لنا ثم نقوم بعد ذلك بوضع آليات الأمان التي تحمينا من هذه الأدوات الجديدة. من هذا المنطلق، تأخذ البيانات الضخمة مكانها إلى جانب مجالات أخرى من المجتمع التي تشكل تحديات لا حل لها، مثل التساؤلات المستمرة عن كيفية تنظيم عالمنا. يجب على كل جيل أن يعالج هذه المشكلات بشكل مستقل، إن مهمتنا هي أن نقدر مخاطر هذه التقنيات القوية وأن ندعم تطورها ــ وأن نحصل على ما تقدمه من مكتسبات.

كما فعلت آلة الطباعة التي قادت إلى تغيير الطريقة التي يحكم بها المجتمع على نفسه، كذلك تفعل البيانات الضخمة. إنها تجبرنا على مواجهة التحديات الجديدة بحلول جديدة. للتأكد من أن الناس محميون وأن التكنولوجيا تتطور في الوقت ذاته، يجب علينا ألا نسمح للبيانات الضخمة بأن تتطور بالقدر الذي يتخطى قدرة البشر على تشكيل هذه التقنية.

10
ما سيحدث لاحقًا

كان "مايك فلاورز"محاميًا بمكتب النائب العام بمقاطعة مانهاتن في أوائل القرن الحادي والعشرين، وكان يعمل على جميع أنواع الجرائم من جرائم القتل إلى الجرائم التي ترتكبها شركات شارع وول ستريت، ثم تحول إلى العمل في شركة قانونية مستقلة. بعد قضاء عام ممل جالسًا خلف مكتبه، قرر أن يترك هذا العمل أيضًا. في أثناء بحثه عن القيام بعمل ذي معنى، فكر في المساعدة في إعادة إعمار العراق. قام أحد شركائه في الشركة بإجراء بعض الاتصالات مع أشخاص مهمين في الدولة. الأمر التالي الذي أدركه "فلاورز"هو أنه كان متوجهًا إلى المنطقة الخضراء، المنطقة الآمنة للقوات الأمريكية في منتصف بغداد، كجزء من الفريق القانوني المشارك في محاكمة "صدام حسين".

تبين بعد ذلك أن أغلب العمل الذي سيقوم به لوجيستيًّا وليس قانونيًّا. كان عليه أن يحدد المناطق التي يشتبه في أنها تحمل مخاطر كبيرة من أجل إرسال المحققين للتحقيق فيها. وكان عليه أن يقوم بهذا الأمر من داخل المنطقة الخضراء دون أن يرسل المحققين ليلقوا حتفهم جراء الهجوم عليهم بالمتفجرات المنتشرة بشكل عشوائي والتي كانت تحدث بشكل يومي ضد القوات الأمريكية. لاحظ أن القوات المسلحة الأمريكية تتعامل مع هذه المهمات على أنها مشكلات معلوماتية. وأن البيانات قد أتت من أجل تقديم يد العون؛ حيث ستقوم أجهزة الاستخبارات بالمزج بين التقارير الميدانية والتفاصيل حول المكان والوقت وخسائر التفجيرات السابقة لتحديد أكثر الطرق الآمنة لتتبعها القوات خلال اليوم.

بعد عودته إلى مدينة نيويورك بعد عدة أعوام، أدرك "فلاورز"أن هذه الأساليب من شأنها أن تكون طرقًا قوية لمكافحة الجريمة أكثر من القوة التي كان يمتلكها في أثناء عمله كمدّعٍ عام. وقد وجد أن عمدة المدينة "مايكل بلومبرج"يمتلك أسلوب التفكير ذاته حيال هذا الأمر، والذي كون ثروته من خلال إمداد البنوك بالمعلومات المالية. دُعي "فلاورز"ليشارك في الفريق الذي سيقوم بمهمة خاصة تتعلق بتحليل الأعداد التي قد تعمل على كشف النقاب عن المخالفات في فضيحة الرهونات عام 2009. حققت هذه الوحدة نجاحًا باهرًا لدرجة أن العمدة "بلومبرج"قد طلب منها أن توسع من مجال عملها في العام التالي. أصبح "فلاورز"أول "مدير للإحصاءات"في المدينة، وكانت مهمته: تكوين فريق من أفضل علماء البيانات الذين يمكنهم إيجادهم من أجل إيجاد وتوظيف معلومات المدينة المهملة لجني الفوائد التي تغطي كل شيء وأي شيء.

ألقى "فلاورز"شبكته بحثًا عن الأشخاص المناسبين؛ حيث قال: "إنني غير مهتم بعلماء الإحصاء المتمرسين؛ حيث إنني أخشى أن يمتنعوا عن اتباع هذه الطريقة الجديدة في علاج المشكلات". قبل هذا، عندما كان يجري مقابلات شخصية مع

أشخاص ذوي سير ذاتية تقليدية من أجل مشروع الاحتيال المالي، كانوا يفصحون عن مخاوفهم من استخدام الأساليب الرياضية؛ حيث قال: "لم أكن أفكر حتى في أيٍّ من النماذج سأقوم باستخدامه، فقد أردت أفكارًا يمكن تنفيذها، وكان هذا كل ما كان يهمني". في النهاية، قام باختيار خمسة أشخاص أطلق عليهم "الأطفال". كانوا جميعًا، ما عدا واحدًا، متخصصين في علم الاقتصاد، تخرجوا في الجامعة منذ عام أو اثنين، لا يملكون الكثير من الخبرة ويعيشون في المدينة الكبيرة، وكانوا جميعًا يمتلكون عقلية ابتكارية.

من بين التحديات الأولى التي واجهت الفريق تحدي "التحويل غير القانوني للمباني"ـــ تقسيم المساكن إلى وحدات أصغر حجمًا بحيث يمكن إيواء عدد أكبر بعشر مرات من العدد الذي صممت من أجله في المقام الأول؛ حيث كانت هذه المساكن عرضة دائمًا للحرائق، كما أنها كانت مستودعات للجريمة والمخدرات والأمراض وانتشار الحشرات الضارة؛ حيث كانت هناك شبكات من تمديدات الأسلاك تزحف على الحوائط؛ وكانت الأوعية الساخنة توضع بشكل خطر على أغطية الأسرة. وكان الأشخاص الذين يقومون بهذه الأمور عادة ما يموتون جراء الحرائق. في عام 2005، لقي اثنان من رجال الإطفاء حتفهم في أثناء محاولتهم إنقاذ القاطنين بهذه المباني. يوجد في مدينة نيويورك وحدها حوالي 25 ألف تحويل غير قانوني للمباني كل عام، ولكنها تمتلك 200 محقق فقط للتعامل معها. لم تكن هناك طريقة جيدة للتفرقة بين الحالات التي تسبب الإزعاج وتلك التي تهدد بنشوب الحرائق. بالنسبة لـ "فلاورز"وأطفاله، كانت هذه المشكلة من المشكلات التي يمكن علاجها بتحليل كم كبير من البيانات.

بدءوا بجمع المعلومات عن جميع الممتلكات الموجودة بالمدينة ـــ والتي بلغ عددها 900 ألف. ثم حصلوا على مجموعات بيانية من 19 وكالة مختلفة تشير إلى، على سبيل المثال، إذا ما كان مالك العقار متهربًا من الضرائب العقارية، إذا ما كانت هناك إجراءات سابقة لإغلاق هذا العقار، أو كانت هناك معدلات غير طبيعية في استهلاك الخدمات أو تم التقصير في دفع المستحقات الحكومية مقابل استخدام تلك الخدمات مما أدى إلى قطعها عن العقار. أضافوا أيضًا معلومات عن نوع المبنى وتاريخ إنشائه، بالإضافة إلى عدد زيارات سيارات الإسعاف للمبنى ومعدلات الجريمة به والشكاوى من وجود القوارض وغيرها. قاموا بعد ذلك بمقارنة جميع هذه المعلومات ببيانات عن الحرائق منذ خمس سنوات مضت مصنفة طبقًا لشدتها وبحثوا عن العلاقات التبادلية من أجل إنشاء نظام قادر على توقع أي من الشكاوى ذات أولوية أعلى.

في البداية، لم تكن أغلب البيانات في الصورة القابلة للاستخدام. على سبيل المثال، لم يستخدم المسئولون عن سجلات المدينة أية طريقة معيارية لوصف الأماكن؛ وبدا أن كل وكالة وإدارة تستخدم أسلوبها الخاص للتعامل مع البيانات، وأعطت إدارة المباني رقمًا فريدًا لكل مبنى. وكان لإدارة الحفاظ على المباني نظام ترقيم مختلف، وأعطت إدارة الضرائب لكل مبنى رقمًا رمزيًّا بناءً على المنطقة الإدارية والحي وقطعة الأرض التي بُني عليها المبنى. واستخدمت الشرطة

نظام الإحداثيات الديكارتية. استخدمت إدارة الإطفاء نظام مدى قرب المبنى من "صناديق الاتصال"بناءً على مكان مركز الإطفاء، رغم أن هذه الصناديق لا تعمل. تعامل أطفال "فلاورز"مع هذه الفوضى من خلال ابتكار نظام يحدد أماكن المباني باستخدام مساحة صغيرة أمام المبنى بناءً على الإحداثيات الديكارتية ثم الاستعانة بالبيانات الجيوجغرافية من قواعد بيانات الوكالات الأخرى. كان أسلوبهم لا يتسم بالدقة، ولكن عوض الكم الهائل من البيانات التي استخدموها عن هذا القصور.

لم يكن أعضاء الفريق سعداء بتحليل الأرقام فقط؛ لذا ذهبوا إلى مواقع الأبنية بصحبة محققين لمراقبة عملهم، ودونوا الكثير من الملاحظات واختبروا وجهات النظر وكل شيء متعلق بالمباني. عندما قال أحد الخبراء وكان كبيرًا في العمر إن المبنى الذي هم بصدد فحصه لن يمثل مشكلة، أراد هؤلاء العباقرة أن يعرفوا سبب ثقته الشديدة تلك، لكنه لم يفصح لهم عن السبب، وبعدها أدركوا بالتدريج أن حدسه اعتمد على أعمال الصيانة الحديثة التي بدت على واجهة المبنى، الأمر الذي جعله يعتقد أن مالك المبنى يهتم به جيدًا.

عاد الأطفال إلى مكاتبهم وهم يتساءلون عن كيفية إضافة "أعمال التجديد الحديثة"إلى نموذج عملهم كنقطة دالة؛ فرغم كل هذا العمل، لم يكونوا قد قاموا برقمنة أعمال التجديد بعد. ولكن لا شك في أن أية أعمال تجديد خارجية تتطلب تصريحًا من المدينة. عند إضافة معلومات هذا التصريح، تحسن الأداء التوقعي لنظامهم بدرجة كبيرة من خلال الإشارة إلى أن بعض المباني المشتبه بها لا تشكل خطرًا كبيرًا.

عادة ما تظهر التحليلات البيانية أن بعض الطرق المثبتة للقيام بالأمور لم تعد هي أفضل الطرق للقيام بها، كما كان على كشافة اللاعبين في فيلم *Moneyball* أن يتقبلوا عيوب استخدام حدسهم في اختيار اللاعبين. على سبيل المثال، كان عدد الاتصال بالخط الساخن لشكاوى المدينة "311" يعتبر هو المؤشر على أي من المباني بحاجة إلى المزيد من الاهتمام، أي أن المزيد من الاتصالات تعني المزيد من المشكلات. ولكن تبين أن هذا المعيار مضلل؛ حيث إن رؤية فأر واحد في الجانب الشمالي الشرقي الأنيق من المدينة قد تسبب في 30 اتصالاً بالشكاوى في ساعة واحدة، ولكن يتطلب الأمر كتيبة من القوارض في حي برونكس حتى يبدأ الناس في التحرك للاتصال برقم "311". وبالمثل، كانت أغلب الاتصالات بشأن التغيير غير القانوني للمنازل تتعلق بالإزعاج، وليس بسبب الحالات التي تعرض العقارات للخطر.

في شهر يونيو من عام 2011، قام "فلاورز" وأطفاله بتشغيل نظامهم؛ حيث كانت جميع الشكاوى التي تصنف على أنها شكاوى متعلقة بالتغيير غير القانوني تعالج أسبوعيًّا. وقاموا بجمع الحالات التي تقع ضمن الخمسة بالمائة الأولى الأكثر عرضة للحرائق وأعطوها للمفتشين من أجل القيام بمتابعة فورية. عندما عاد المفتشون بالنتائج، كانت نتائج مذهلة للجميع.

قبل تحليل البيانات الضخمة، كان المفتشون يتابعون الشكاوى التي بدت أكثر خطورة، ولكن كانت نسبة 13 % فقط هي التي وجدوا أن حالتها تستدعي إصدار

أمر بإخلائها، ولكن في الوقت الحالي، أصبحوا يصدرون أوامر بإخلاء 70% من المباني التي يقومون بفحصها. من خلال تحديد أي من المباني تحتاج إلى المزيد من الاهتمام، حسنت البيانات الضخمة من فاعليتهم خمس مرات. وأصبح عملهم مرضيًا أكثر من ذي قبل: أصبحوا يركزون على المشكلات الأضخم. إن الفاعلية الجديدة للمفتشين لها عدد لا يحصى من الفوائد؛ حيث أصبحت الحرائق في قطاع التغيير غير القانوني أفضل بخمس عشر مرة فيما يتعلق بموت أو إصابة رجال الإطفاء من الحرائق الأخرى؛ لذا فإن إدارة الحماية المدنية أحبت هذه التحليلات. بدا "فلاورز" وأطفاله مثل السحرة الذين يملكون كرة سحرية تمكنهم من التنبؤ بالمستقبل وتخبرهم بأي الأماكن أكثر عرضة للمخاطر. لقد استغلوا كميات هائلة من البيانات التي كانت مهملة طوال سنوات؛ حيث كانت تقبع دون استخدام بعد جمعها، ووظفوها بطريقة جديدة لاستخراج القيمة الحقيقية منها. إن استخدام كميات كبيرة من المعلومات سمح لهم باكتشاف العلاقات التي لم تكن ظاهرة في كميات المعلومات الأصغر حجمًا ــ جوهر البيانات الضخمة.

إن التجربة التي أجراها علماء تحليل البيانات في مدينة نيويورك، ألقت الضوء على الكثير من الموضوعات العامة التي تناولها هذا الكتاب؛ حيث استخدموا كميات هائلة من البيانات، وليس بعضها فقط، وعبرت قوائمهم بمباني المدينة على أنهم قد استخدموا معادلة ن = الجميع. كانت البيانات فوضوية، مثل معلومات الأماكن وسجلات سيارات الإسعاف، ولكن هذا لم يثبط من عزيمتهم. في حقيقة الأمر، كانت فوائد استخدام المزيد من البيانات تتخطى عيوب قلة المعلومات الدقيقة. وأصبحوا قادرين على تحقيق ما يأملون بتحقيقه لأن الكثير من ملامح المدينة قد تمت رقمنتها (ولكن بشكل غير ثابت)؛ مما سمح لهم بمعالجة المعلومات.

يجب أن تتنحى المعرفة المحدودة للخبراء عن الطريق وتسمح للأساليب القائمة على تحليل البيانات بالمرور. وفي الوقت ذاته، كان "فلاورز" وأطفاله يواصلون اختبار نظامهم بالتعاون مع الخبراء المتمرسين مستقين منهم خبراتهم من أجل تحسين أداء النظام. ولكن السبب الرئيسي في نجاح هذا النظام هو أنه تخلى عن تقصي الأسباب في سبيل الحصول على العلاقات التبادلية.

يوضح "فلاورز": "أنا لا أهتم بالأسباب إلا إن كانت تؤدي إلى حدوث أمر ما. إن الأسباب وجدت من أجل أشخاص آخرين، وبصراحة إن الأمر يحمل خطورة كبيرة عندما تبدأ الحديث عن الأسباب. إنني لا أعتقد أنه يوجد أي سبب بين اليوم الذي يصدر فيه شخص ما أمرًا بإغلاق لبناية وما إذا كان هذا المكان يمتلك تاريخًا في حرائق الأبنية. أعتقد أنه من الغباء أن نفكر بهذه الطريقة. ولن يتجرأ أي شخص على الخروج وقول هذا الكلام. سوف يفكرون، لا، إنها العوامل الخفية. ولكنني لا أرغب حتى في التطرق لهذا الأمر. إنني بحاجة إلى نقاط بيانية محددة يمكنني الحصول عليها، وهي من سيخبرني بأهمية الأمر، فإن كان مهمًّا، فسوف نعمل عليه. أما إن لم يكن مهمًّا، فإننا لن نعمل عليه. إنك تعلم أننا نواجه مشكلات حقيقية لنعالجها. لا يمكنني أن أهيم عابثًا دون هدف، بصراحة، وأفكر في أمور أخرى مثل أسباب ما يحدث".

عندما تتحدث البيانات

إن للبيانات الضخمة تأثيرًا كبيرًا على المستوى العملي؛ حيث إنه يتم تطبيق هذه التكنولوجيا لإيجاد حلول للمشكلات اليومية المزعجة. ولكن هذه مجرد البداية. تعمل البيانات الضخمة على إعادة تشكيل أسلوب حياتنا وعملنا وتفكيرنا. إن التغيير الذي نشهده في بعض المجالات يكون أعظم من المبتكرات العصرية السابقة التي تسببت في توسيع جذري لمجال استخدام المعلومات في المجتمع. إن الأرض من تحت أقدامنا تتغير، وأصبحنا نتشكك في المعتقدات القديمة. تتطلب البيانات الضخمة مناقشات جديدة عن طبيعة اتخاذ القرارات والقدر والعدالة. إن وجهة نظرنا عن العالم على أنه مصنوع من الأسباب قد تم تحديها بواسطة العلاقات التبادلية التي تفوقت عليها. أصبح امتلاك المعرفة، الذي كان يعني في الماضي فهم الماضي، يعني في الوقت الحالي القدرة على توقع المستقبل.

أصبحت هذه الموضوعات أكثر أهمية من الموضوعات التي عبرت عن نفسها عندما استعددنا لاستخدام التجارة الإلكترونية، والتعايش مع الإنترنت، ودخول عصر الحواسب، أو استخدام المعداد. إن الفكرة من إضفاء أهمية كبرى على سعينا إلى فهم الأسباب تكون أكبر مما تستحق؛ حيث إنه في الكثير من الأحيان يكون من المفيد أكثر أن نتجنب السؤال عن السبب لصالح النتائج ــ تعني أن هذه الأمور مهمة لمجتمعنا ولوجودنا. إن التحديات التي فرضتها البيانات الضخمة ربما لا يمكنها أن تنتج إجابات. ولكنها تعتبر جزءًا من الجدل الأبدي عن مكان الإنسان من الكون وبحثه عن معنى في خضم ضجة هذا العالم الفوضوي الذي لا يمكن فهمه.

وفي النهاية، أصبحت البيانات الضخمة حدث الساعة عندما أوفى "مجتمع المعلومات"بوعده الذي قطعه على نفسه، أن تصل البيانات للصدارة، أصبح من الممكن في الوقت الحالي أن نستخدم جميع هذه الملفات الرقمية التي جمعناها بطرق جديدة لخدمة أهداف جديدة ولاكتشاف أنواع جديدة من القيم، ولكن يتطلب هذا الأمر طرقًا جديدة للتفكير وسوف تتحدى مؤسساتنا وحتى حسنا بالهوية. الأمر الوحيد المؤكد هو أن كمية البيانات ستواصل النمو، وكذلك القدرة على معالجتها بالكامل. ولكن عندما يعتبر أغلب الناس أن البيانات الضخمة موضوع تقنيّ مع التركيز على البرمجيات أو أجهزة الحاسب، فإننا نؤمن بأن التركيز يجب أن يتحول إلى ما يحدث عندما تتحدث البيانات.

يمكننا أن نجمع ونحلل كمًّا من المعلومات أكبر مما كان من قبل. لم تعد ندرة البيانات, هي السمة التي تحدد جهودنا لفهم العالم. يمكننا أن نوظف المزيد من البيانات وفي بعض الأحيان، يمكننا الاقتراب منها جميعًا، ولكن القيام بهذا الأمر يجبرنا على العمل بطرق غير تقليدية، وبالأخص، تغيير أفكارنا عن ماهية المعلومات المفيدة.

بدلاً من أن نشغل أنفسنا بدقة وصحة وتنظيم وقدرة البيانات، يمكننا أن نسمح ببعض الفوضى. يجب علينا ألا نقبل البيانات على أنها صحيحة أم خطأ، ولكن قد نقبل بعض الفوضى في مقابل الحصول على كم أكبر من المجموعات البيانية

الشاملة. في الواقع، في بعض الأحيان تكون البيانات الضخمة والفوضوية مفيدة للغاية؛ حيث إننا كنا معتادين استخدام كميات ضئيلة منها، الجزء الدقيق من البيانات، وانتهى بنا الأمر بالفشل في الحصول على كم التفاصيل الذي يوجد به القدر الأكبر من المعرفة.

بسبب أن العلاقات التبادلية يمكن إيجادها بصورة أسرع وأرخص من الأسباب، فعادة ما نفضلها على الأسباب. سنظل بحاجة إلى الدراسات السببية والتجارب المخطط لها إلى جانب البيانات المعالجة بعناية في حالات بعينها، مثل تصميم قطعة غيار مهمة للطائرات. ولكن بالنسبة للكثير من الاحتياجات اليومية، تعتبر معرفة النتائج بدلاً من الأسباب أمرًا كافيًا. ويمكن للبيانات الضخمة أن تحدد الطريق نحو مجالات واعدة التي يمكننا من خلالها استكشاف العلاقات السببية.

تسمح لنا هذه العلاقات التبادلية السريعة بأن نوفر المال عند شراء تذاكر الطيران، وتوقع انتشار أوبئة الأنفلونزا، وتحديد أي من فتحات الصرف أو البنايات المزدحمة التي تجب علينا متابعتها في هذا العالم المحدود المصادر. قد تمكن شركات التأمين الصحي من أن تمد الناس بالتغطية التأمينية دون الحاجة إلى إجراء فحوصات طبية، وتقلل تكليف تذكير المرضى بتناول عقاقيرهم. أصبحت اللغات تُترجم وأصبحت السيارات تقود نفسها على أساس التوقعات التي تم الحصول عليها من خلال العلاقات التبادلية للبيانات الضخمة. تمكنت شركة وولمارت من معرفة أية نكهة لمنتج بوب تارت عليها وضعها في واجهة المتاجر في أثناء الأعاصير. (الإجابة هي: الفراولة). لا شك في أن معرفة الأسباب أمر جيد إذا ما تمكنت من اكتشافها. المشكلة في أنها عادة ما يكون من الصعب اكتشافها، وعندما نظن أننا قد اكتشفناها، فإننا عادة ما نخدع أنفسنا.

الأدوات الجديدة، من المعالجات الأكثر سرعة والذاكرات الكبرى سعة إلى البرامج والخوارزميات الأكثر ذكاءً، لا تشكل إلا جزءًا من الأسباب التي تمكننا من القيام بجميع هذه الأمور. رغم أن الأدوات مهمة، فإن السبب الرئيسي في تحقيق كل هذه الأمور هو أننا نمتلك المزيد من البيانات؛ حيث إنه تمت رقمنة المزيد من سمات العالم من حولنا. الأمر الأكيد، أن طموح البشر لقياس العالم قد سبق ثورة الحواسب بوقت طويل. ولكن قامت الأدوات الرقمية بتسهيل الرقمنة بشكل كبير؛ حيث لم تتمكن الهواتف المحمولة مِنْ تتبع مَنْ يجري المكالمات وأين نذهب فحسب، بل يمكن استخدام البيانات التي جمعتها لتحديد ما إذا كنا سنتعرض للإصابة بالمرض. في وقت قريب، قد تتمكن البيانات الضخمة من توقع ما إذا كنا سنصاب بالمرض.

إن قدرتنا على القيام بالأمور الجديدة والإضافية والأفضل والأسرع تمكننا من فك قيود قيم ضخمة، وأن نوجد أنواعًا جديدة من الفائزين والخاسرين. تأتي أغلب قيمة البيانات من استخداماتها الثانوية، وقيمتها الاختيارية، وليس من استخدامها الأولي، مثلما اعتدنا أن نفكر فيها من قبل. نتيجة لهذا، بالنسبة لأغلب أنواع البيانات، من المنطقي أن نجمع أكبر كم يمكننا جمعه وأن نحتفظ بها طالما أنها تصدر قيمًا جديدة، وأن ندع الآخرين يحللونها إذا ما كانوا أفضل منا في استخراج القيم منها

(شريطة أن يشارك المرء المكاسب التي تنتج عنها).

إن الشركات التي يمكنها أن تضع نفسها في منتصف تدفق المعلومات ويمكنها جمع البيانات ستزدهر. إن توظيف البيانات الضخمة بفاعلية يتطلب مهارات تقنية والكثير من الخيال ــ عقلية البيانات الضخمة. ولكن قد تحصل الشركات المالكة للبيانات على أغلب القيمة. وأحيانًا ما ستكون الأصول المهمة أمرًا يتعدى مجرد المعلومات الواضحة بل أيضًا بقايا البيانات التي تنشأ عن تعاملات الناس مع المعلومات، التي يمكن للشركات الماهرة أن تستخدمها من أجل تحسين خدماتها الحالية أو لإطلاق خدمات جديدة.

في الوقت ذاته، تضع أمامنا البيانات الضخمة مخاطر ضخمة؛ حيث إنها كشفت الآليات التقنية والقانونية الجوهرية غير الفعالة التي نعمل من خلالها على حماية الخصوصية. في الماضي، كان ما يُكوِّن المعلومات التي يمكن من خلالها تحديد الأشخاص معروفًا للجميع ــ الأسماء وأرقام التأمين الاجتماعي وسجلات الضرائب وغيرها ــ وكان من السهل حمايتها نسبيًّا. أما الآن فحتى البيانات غير المؤذية يمكنها كشف هوية الأشخاص إذا ما تمكن جامع البيانات من جمع كَمٍّ كافٍ منها. أصبح التجهيل أو الإخفاء عن الأعين غير صالح للعمل. هذا بالإضافة إلى أن استهداف الأشخاص بالمراقبة أصبح يتضمن في الوقت الحالي المزيد من الاعتداء السافر على الخصوصية أكثر من أي وقت مضى؛ حيث إن السلطات لم تعد راغبة في معرفة أكبر كم ممكن من المعلومات عن الأفراد فحسب، بل أيضًا أكبر عدد من العلاقات والمعارف والتعاملات.

إلى جانب تحدي الخصوصية، كشفت استخدامات البيانات الضخمة هذه عن مخاوف مزعجة وفريدة من نوعها: المخاطرة بأننا قد نحكم على الناس ليس فقط من خلال سلوكياتهم الواقعية بل من خلال توجهاتهم التي افترضت البيانات أنهم يمتلكونها. مع زيادة دقة توقعات البيانات الضخمة، يمكن أن يستخدمها المجتمع لعقاب الناس على سلوكياتهم المتوقعة ــ الأفعال التي لم يقوموا بها بعد. التوقعات التي لا يمكن إثبات خطئها؛ حيث لن يمكن للأشخاص الذين تم اتهامهم بأن يبرئوا أنفسهم. إن العقاب على هذا الأساس يلغي مفاهيم الإرادة الحرة والاحتمالية، مهما كانت محدودة، بأن المرء قد يختار تصرفًا آخر. مع تحديد المجتمع للمسئولية الفردية (ويأمر بالعقاب وفقًا لها)، يجب اعتبار الإرادة البشرية منيعة. يجب أن يبقى المستقبل أمرًا يمكننا تشكيله طبقًا لتصميمنا الخاص؛ لأنه إن لم يكن كذلك، فإن البيانات الضخمة ستقوم بتشويه جوهر الإنسانية: التفكير المنطقي والإرادة الحرة.

لا توجد طرق مثبتة للاستعداد التام لعالم البيانات الضخمة؛ حيث إنها تطلب إنشاء مبادئ جديدة نحكم على أنفسنا من خلالها. قد تساعد سلسلة من التغييرات المهمة لممارساتنا المجتمع على أن يصبح أكثر معرفة لشخصية وعيوب البيانات الضخمة. يجب علينا أن نحمي خصوصيتنا من خلال تحويل المسئولية بعيدًا عن الأشخاص لنلقي بها على عاتق مستخدمي البيانات ــ من أجل الاستخدام الحسابي. في عالم التوقعات، من المهم أن نتأكد من أن تظل الإرادة البشرية في مكان مقدس ولا

يجب أن نحافظ فقط على قدرة الناس على الخيار الأخلاقي، بل أيضًا على المسئولية الفردية للإقدام على التصرف. على المجتمع أيضًا أن يصمم إجراءات احترازية لتسمح للفئات المهنية الجديدة من "متخصصي الخوارزميات"بأن يقيموا تحليلات البيانات الضخمة؛ بحيث لا يتحول العالم الذي قلت عشوائيته بفضل البيانات الضخمة إلى صندوق أسود، أي لا نستبدل بأحد الأمور المجهولة أمورًا أخرى.

ستصبح البيانات الضخمة جزءًا لا يتجزأ من استيعاب وعلاج الكثير من المشكلات البيئية الملحة في عالمنا؛ حيث يتطلب علاج مشكلات المناخ تحليل بيانات التلوث من أجل تحديد المجالات التي علينا أن نركز عليها بشكل أكبر وأن نوجد طرقًا للحد من هذه المشكلات. لقد تم وضع حساسات في جميع أنحاء العالم، بما فيها تلك المدمجة في الهواتف الذكية، والتي تمدنا بكم هائل من البيانات التي من شأنها أن تمكننا من دراسة الاحتباس الحراري العالمي بشكل أكثر تفصيلاً. في الوقت ذاته، فإن تقليل تكاليف الرعاية الصحية، خاصة بالنسبة للفقراء، سيحدث من خلال التشغيل الآلي للمهام التي تبدو في الوقت الحالي بحاجة إلى الحكم البشري ولكن يمكن القيام بها بواسطة الحاسب؛ مثل فحص الأنسجة للخلايا السرطانية أو تحديد أماكن العدوى قبل ظهور الأعراض.

لقد بدأ بالفعل استخدام البيانات الضخمة في تحسين الاقتصاد ولمنع النزاعات؛ حيث إنها كشفت عن العديد من الأحياء الفقيرة في إفريقيا التي تعتبر مجتمعات حيوية تصلح للنشاط الاقتصادي من خلال تحليل تحركات مستخدمي الهواتف المحمولة. كما كشفت عن المناطق المعرضة للنزاعات الطائفية وأشارت إلى الكيفية التي يمكن من خلالها حل مشكلات اللاجئين. وسوف تتضاعف استخداماتها مع تطبيق هذه التقنية على أوجه أخرى من حياتنا.

تساعدنا البيانات الضخمة على القيام بما نقوم به بالفعل ولكن بشكل أفضل، كما تسمح لنا بأن نقوم بأمور جديدة تمامًا. ولكن رغم هذا فهي ليست عصا سحرية؛ حيث إنها لن تعمم السلام على العالم، أو تقضي على الفقر، أو يظهر من خلالها "بيكاسو"جديد. لا يمكن للبيانات الضخمة أن تلد طفلاً ــ ولكنها قادرة على إنقاذ المبتسرين منهم. في وقت ما، سنتوقع منها أن تُستخدم في جميع مجالات الحياة تقريبًا، وربما نشعر بالقلق إذا ما غابت عنا، بالطريقة ذاتها التي نتوقع من الطبيب أن يجري فحصًا بالأشعة السينية ليكتشف المشكلات التي لم يتمكن من اكتشافها بالفحص التقليدي.

مع اعتياد الناس وجود البيانات الضخمة، قد تؤثر على أسلوب تفكيرنا بالمستقبل. منذ حوالي خمسمائة عام، مرت البشرية بتحول كبير في فهمها للزمن، كجزء من تحول القارة الأوربية إلى قارة أكثر دنيوية وتركيزًا على العلم وتنورًا. قبل هذا، كان ينظر إلى الوقت على أنه دورة وكذلك كانت الحياة؛ حيث كان كل يوم (وعام) يشبه اليوم الذي سبقه، حتى إن نهاية الحياة تؤدي إلى بدايتها؛ حيث إن البالغين يتحولون إلى أطفال مرة أخرى. بعد ذلك أصبح يُنظر للزمن على أنه خطي ــ تتابع لا نهائي

من الأيام التي يتشكل العالم من خلالها ويتأثر أيضًا مسار الحياة. في الماضي، كان الماضي والحاضر والمستقبل ممزوجين معًا، أما الآن فقد أصبح للإنسانية ماض لتتذكره، ومستقبل لتتطلع إليه؛ حيث إنها قد شكلت حاضرها.

رغم أنه يمكن صياغة الحاضر، فإنه قد تم تحويل المستقبل من أمر قابل للتوقع إلى أمر واضح وأصيل ــ لوحات كبيرة وخالية يمكن للمرء أن يرسم عليها ما يتوافق مع قيمه وجهوده. من بين السمات المحددة للعصر الحديث هو رؤيتنا أنفسنا على أننا مسيطرون على حياتنا، وهذا التوجه هو ما يميزنا عن أسلافنا، الذين كانت الحتمية بالنسبة لهم هي الأمر الطبيعي. ولكن تعيد توقعات البيانات الضخمة المستقبل إلى حالته الأولى من عدم الوضوح والغموض. وبدلاً من أن يكون مجرد لوحات خالية، أصبح مستقبلنا مرسومًا بخطوط باهتة لا يميزها سوى من يملكون التقنيات اللازمة لتوضيحها. من شأن هذا الأمر أن يقلل من قدرتنا على التحكم في قرارتنا، والتي ربما قد تمت التضحية بها على مذبح الاحتمالية.

في الوقت ذاته، قد تعني البيانات الضخمة أن نظل سجناء للأبد جراء أفعالنا السابقة، والتي يمكن استخدامها ضد مصلحتنا في الأنظمة التي تزعم أنها تتوقع السلوك المستقبلي: لا يمكننا الهرب مما اقترفته أيدينا من قبل. كتب "شكسبير": "إن ما مضى هو التمهيد لما سيحدث". إن البيانات الضخمة تحتفظ بهذا في الصورة الخوارزمية، بالنسبة للأمور الجيدة والسيئة على حد سواء. هل سيعمل العالم القائم على التوقعات على أن يثبط من فرحتنا بشروق الشمس، ورغبتنا في أن نضع بصمتنا البشرية في هذا العالم؟

في حقيقة الأمر، النقيض هو الأكثر أرجحية. إن معرفة ما يمكن للأفعال أن تحدثه في المستقبل ستسمح لنا بأن نقدم على خطوات إصلاحية لمنع تفاقم المشكلات أو لتحسين النتائج. سيمكننا أن نحدد الطلبة المهددين بالفشل قبل الامتحانات بوقت طويل. سنكون قادرين على تحديد الخلايا السرطانية وعلاجها قبل أن ينتشر المرض. سيمكننا رؤية احتمالية حدوث الحمل غير الشرعي أو الانجراف في طريق الجريمة ويمكننا التدخل للحيلولة دون الوصول إلى هذه النتائج المتوقعة، بقدر إمكاننا. سنكون قادرين على منع الحرائق المميتة عن التهام مساكن مدينة نيويورك المزدحمة بالسكان من خلال معرفة أي الأبنية علينا أن نفحصها.

إن البيانات الضخمة قابلة للتغيير؛ حيث إنها مجرد نتائج محتملة، وهذا يعني أنه إن أردنا أن نغيرها فإننا قادرون على ذلك بالسعي والعمل الدءوب. يمكننا أن نحدد أفضل الطرق لاستشراف المستقبل، مثلما تمكن "موراي"من إيجاد الطرق الطبيعية في هذا الخضم الهائل من الرياح والأمواج.

البيانات الأضخم

مع تغيير البيانات الضخمة لحياتنا ــ من خلال تعديلها وتحسينها وجعلها أكثر فاعلية والحصول على أكبر فائدة منها ــ ما الدور الذي تركته ليلعبه الحدس والإيمان وعدم الثقة والأصالة؟

إذا كانت البيانات الضخمة قد علمتنا أي شيء، وهي أن التصرف بشكل أفضل

وإحداث التحسينات ــ دون مزيد من الفهم ــ أمور كافية. عادة ما يكون القيام بهذه الأمور فعالاً. حتى إن لم تعلم سبب نجاح جهودك، إنك تنتج نتائج أفضل مما كنت تفعل قبل بذل مثل هذه الجهود. قد لا يجسد "فلاورز"وأطفاله المعرفة التي يتميز بها الحكماء، ولكنهم تمكنوا من إنقاذ حياة الكثيرين.

إن البيانات الضخمة ليست مجرد عالم بارد من الخوارزميات والآلات. هناك دور رئيسي للبشر، مع كل نقاط ضعفنا ومعتقداتنا المضللة وأخطائنا؛ حيث إن هذه الصفات تصاحب الابتكار والغريزة والعبقرية البشرية. إن العمليات العقلية الفوضوية ذاتها التي تؤدي بنا إلى المهانة أو إساءة الحكم هي ما يعطي دفعة للنجاحات والوصول إلى العظمة. من هذا المنطلق، مثلما تعلمنا تقبل البيانات الفوضوية لأنها تخدم هدفًا أسمى، فعلينا أن نتقبل عدم الدقة التي تعتبر من سمات البشر؛ ففي نهاية الأمر، تعتبر الفوضى جزءًا أساسيًّا من العالم ومن عقولنا، في كلتا الحالتين، لن نستفيد إلا إذا تقبلناها وطبقناها.

في العالم الذي تحدد فيه البيانات اتخاذ القرارات، فما الهدف الذي سيبقى ليحيا من أجله البشر، أو ماذا سيكون الغرض من الحدس ومعارضة الحقائق؟ إذا ما أعجبت البيانات الجميع وتمكن الجميع من توظيف أدوات البيانات الضخمة، فربما ما سيصبح النقطة الرئيسية الفارقة هو عدم القدرة على التوقع: قدرات البشر الغريزية، الإقدام على المخاطر، الحوادث والأخطاء.

إذا كان الأمر كذلك، فستكون هناك حاجة ملحة لتحديد مكان للبشر: من أجل الحفاظ على مساحة للحدس والمنطق السليم والمواهب للتأكد من أنها غير مزدحمة بالبيانات والإجابات التي أجابت بها الآلات. إن الأمور العظيمة الموجودة داخل البشر هي الأمور التي لا يمكن للخوارزميات أو لرقاقات السيليكون كشفها، لا يمكنها كشفها لأنها لا يمكن أن تتحول إلى بيانات. إنها ليست "الأمر"بل إنها "عدم حدوث الأمر": المساحات الخالية والشقوق في الأرصفة والكلام الذي لم يُقَلْ والأفكار التي لم يتم التفكير فيها بعد.

إن لهذا الأمر مقتضيات مهمة فيما يتعلق بفكرة التقدم في المجتمع؛ حيث تمكننا البيانات الضخمة من إجراء التجارب بصورة أسرع وأن نختبر المزيد من الأدلة. من شأن هذه المزايا أن ينتج عنها المزيد من الابتكارات. ولكن تأتي شرارة الاختراعات مما لا تفصح عنه البيانات. هذا الأمر لا يمكن لأية كمية من البيانات أن تؤكده أو تثبته؛ حيث إنها غير موجودة في الواقع. إذا ما كان "هنري فورد"بالتنقيب في خوارزميات البيانات الضخمة من أجل معرفة ما يحتاج إليه عملاؤه، لكانت أجابت عليه: "حصان أسرع". (من أجل إعادة صياغة مقولته الشهيرة). في عالم البيانات الضخمة، إن أكثر صفاتنا تعبيرًا عن الإنسانية هي ما يجب أن يتم غرسها فينا ــ ابتكارنا وحدسنا وطموحنا الفكري ــ حيث إن عدم أصالتنا هي مصدر تقدمنا.

إن البيانات الضخمة عبارة عن مصدر وأداة؛ حيث إنها قد أنشئت من أجل الإخبار وليس التوضيح، إنها توجهنا نحو الفهم، ولكنها أيضًا قد تؤدي بنا إلى سوء الفهم، بناءً على مدى جودة أو سوء استخدامها. ومهما انبهرنا بقوة البيانات الضخمة، فيجب علينا ألا ندع رونقها المغري أن يعمينا عن عيوبها المتأصلة.

إن جملة المعلومات في العالم ــ المعادلة المطلقة ن = الجميع ــ لا يمكن جمعها أو تخزينها أو معالجتها بواسطة تقنياتنا الحالية. على سبيل المثال، معمل الفيزياء الجزيئية "سيرن"بسويسرا يقوم بجمع ما يقل عن 0.1% من المعلومات التي تنتج عن تجاربه ــ يبدو أن البقية التي لا فائدة منها، تُترك لتتبدد في الأثير. ولكن يعتبر هذا الأمر حقيقة جديدة بالكاد. إن المجتمع عادة ما يكون متعلقًا بقيود الأدوات التي نستخدمها لقياس وإدراك الواقع، من البوصلة والآلة السدسية إلى المنظار والرادار إلى نظام تحديد المواقع العالمي في الوقت الحاضر. إن أدواتنا قد تصبح أقوى بضعف أو عشرة أضعاف أو ألف ضعف مما هي عليه الآن، وهو ما يجعل الأدوات التي نستخدمها في الوقت الحاضر صغيرة ومحدودة. سيبدو عالم البيانات الضخمة الحالي، ليس بعد وقت طويل، غريبًا كما أصبحت الذاكرة القابلة للكتابة ذات الأربعة كيلوبايت على متن مركبة الفضاء أبوللو 11 المسئولة عن توجيه السفينة بالنسبة لنا في الوقت الحاضر.

إن ما سيمكننا أن نجمعه ونعالجه سيظل جزءًا يسيرًا من المعلومات الموجودة في العالم. قد تكون محاكاة تافهة للواقع، مثل الخيالات المرسومة على جدران كهف أفلاطون؛ لأننا لن نستطيع أبدًا الحصول على معلومات مثالية، كما أن توقعاتنا ستظل ضعيفة. لا يعني هذا أنها خطأ، بل إنها دائمًا ما ستكون ناقصة. إنها لن تلغي الأفكار التي تقدمها البيانات الضخمة، ولكنها تضع البيانات الضخمة في مكانها الحقيقي ــ الأداة التي لا يمكنها تقديم الإجابة الشافية، ولكنها أداة كافية لمساعدتنا في الوقت الحالي حتى تظهر أساليب أفضل وبالتالي إجابات أفضل في المستقبل. يقترح هذا الأمر أيضًا أنه علينا أن نستخدم هذه الأداة بأكبر قدر من التواضع... والإنسانية.

ملاحظات

1. الآن

1. نماذج توقع انتشار الإنفلونزا من شركة جوجل - جيريمي جينسبرج وآخرين
"Detecting
Influenza Epidemics Using Search Engine Query Data," Nature 457 (2009), pp. 1012-14 (http://www.nature.com/nature/journal/v457
/n7232/full/nature07634
.(.html

2. متابعة لدراسة عن نماذج توقع انتشار الإنفلونزا من جوجل - إيه إف دوجاس وآخرين،
"Google
Flu Trends: Correlation with Emergency Department Influenza Rates and Crowding Metrics" CID Advanced Access (January 8, 2012); DOl .10.1093/cid/cir883

3. شراء تذاكر الطيران, فاير كاست - حصلنا على المعلومات من كينيث كوكيير،
"Data,
Data Everywhere;' The Economist special report, February 27,2010, pp.1-14,
ومن المقابلات الصحفية مع إتزيوني ما بين عامي 2010 و2012. Etzioni's Hamlet project - أورين إتزيوني، وسي إيه نوبلوك، وآر توشيندا، وإيه ييتس،
"To
Buy or Not to Buy : Mining Airfare Data to Minimize Ticket Purchase Price" SIGKDD '03, August 24-27, 2003 (http://knight.cis.tem ple.edu/-yates//papers/hamlet-
.(kdd03.pdf

4. الثمن الذي دفعته مايكروسوفت لشراء فاير كاست - من تقارير وسائل الإعلام،
"Secret
Farecast Buyer Is Microsoft," Seattlepi.com, April 17, 2008 (http://blog.seattlepi.com/venture/2008/04/17 /secret-farecast-buyer-is-microsoft/?
.(source=mypi

5. طريقة واحدة للتفكير في البيانات الضخمة - هناك جدل محتدم وغير مثمر على أصل مصطلح "البيانات الضخمة"، وعما إذا كان يعبر عنها بالفعل أم لا. فقد تم استخدام هاتين الكلمتين معًا بتناغم لفترة طويلة من الزمن. حددت دراسة بحثية أجراها دوج لاني عام 2011، السمات الثلاث للبيانات الضخمة (الحجم، السرعة والتنوع)، والتي كانت مفيدة للغاية في حينها ولكنها غير كاملة.

6. الفلك وتسلسل الحمض النووي - كوكيير, "Data, Data Everywhere." Billions of shares traded ريتا - نازارس وجوليا ليتي "Stock, Trading in U.S. Falls to Lowest Level Since 2008," Bloomberg, August 13,2012 (http://www.bloomberg.com/news/2012-08-13/stock-trading-in-u-s-hits-lowest-Ievel-since-2008-as-vix-falls.html).

7. 24 بيتا بايت من شركة جوجل يوميًّا - توماس إتش دافنبورت، بول بارث وراندي بين، "How 'Big Data' Is Different" Sloan Review, July 30, 2012, pp. 43-46 (http://sloanreview.mit.edu/the-magazine/2012-fall/54104/how-big-data-is-different/).

إحصائيات فيس بوك - دعاوى الاكتتاب العام لشركة فيس بوك، "Form S-l Registration State ment:' U.S. Securities and Exchange Commission, February 1, 2012 (http://sec.gov / Archives/edgar /data/1326801/000119312512034517 /d287954dsl.htm)

إحصاءات يوتيوب - لاري بايج، "Update from the CEO" Google, April 2012 (http://investor.google.com/corporate/2012/ceo-Ietter.html).

عدد التغريدات - توميو جيرون, "Twitter's Dick Costolo: Twitter MobileAd Revenue Beats Desktop on Some Days,"Forbes, June 6, 2012 (http://www.forbes.com/sites/tomiogeron/2012/06/06/twitters-dick-costolo-mobile-ad-revenue-beats-desktop-on-some-

.(/days

8. معلومات عن كمية البيانات - مارتن هلبرت وبريسيلا لوبيز، The“ World’s Technological Capacity to Store, Communicate, and Com pute Information” Science, April, 20ll, pp.60-65;مارتن هلبرت وبريسيلا لوبيز، How“ to Measure the World’s Technological Capacity to Communicate, Store and Compute Information?” International Journal of Communication 2012, pp. 1042-55 (http://www.ijoc.org/ojs/index.php/ .(ijoc/article/viewFile/1562/742

9. تقدير لكمية المعلومات المخزنة عام 2013 - مقابلة كوكيير مع هلبرت, 2012.

10. آلة الطباعة والثمانية ملايين كتاب؛ الأكثر إنتاجًا منذ تأسيس القسطنطينية - إليزابيث إل. إينشتين, The Printing Revolu tion in Early Modern Europe (Canto/Cambridge University Press, 1993), pp.13-.14

تشبيه بيتر نورفيج - من حوارات نورفيج مع الصحيفة: إيه هالفي وبيتر نورفيج وإف بيريرا، The“ Unreasonable Effectiveness of Data” IEEE Intelligent Systems, March/April 2009, pp. 8-12 mputer.org/portal/cms_docs_intelligent/intelligent/homepage/2009/x2exp.pdf). (Note that the title is a play on Eugene Wigner’s article “The Unrea- sonable Effectiveness of Mathematics in the Natural Sciences” in which he considers why physics can be nicely expressed in basic math but the social sciences resist such tidy .formulas

انظر مقال سي ويجنر، The“ Unreasonable Effectiveness of Mathematics in the Natural Sciences,” Communications on Pure and Applied Mathematics 13, no. 1 (1960), pp. 1-14.) Among Norvig’stalks on the paper is “Peter Norvig-The Unreasonable Effectiveness of Data,” lecture at University of British Columbia, YouTube, September 23,2010 (http://www.youtube.com/watch?

.(v=yvDCzhbjYWs

11. عن الحجم المادي المؤثر على القانون المادي المعمول به (رغم أنه غير صحيح بالكامل)، المرجع الرئيسي عن هذا الأمر هو جاي بي إس هالدان، On“
Be ing the Right Size” Harper’s Magazine, March 1926 (http://harpers.org /archive/1926/03/on-being-the-right-
.(/size

12. بيكاسو في صور لاسكو - دافيد وايتهاوس، UK“
Science Shows Cave Art Developed Early” BBC News Online, October 3, 2001 (http://
.(news.bbc.co.uk/1/hi/sci/tech/1577421.stm

2. المزيد

1. اقتباسات جيف جوناس - حوار مع جوناس، December 2010, .Paris

2. تاريخ الإحصاء السكاني بالولايات المتحدة الأمريكية - مكتب الإحصاء الأمريكي، The“
Hollerith Machine” Online history. (http://www.census.gov/history/www/innovations/tech nology/the hollerith .tabulator.html

3. مساهمات نيمان - وليام كروسكال وفريدريك موستيللر، Representative“
Sampling, IV: The History of the Concept in Statistics, 1895-1939,”International Statistical Review 48 (1980), pp. 169-195, pp. 187-188. Neyman’s famous paper is Jerzy Neyman, “On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection:’ Journal of the Royal Statistical Society 97, no. 4 (1934),pp. 558-
.625

العينة المكونة من ألف ومائة شخص كافية - إيرل بابي، Practice of Social Research (12th ed. 2010), pp. 204-
.207

4. تأثير الهواتف الخلوية —

Estimating“
the Cellphone Effect:’ September 20,2008 (http://www.fivethirtyeight.com/2008 /09 /estimating-cellphone-effect-22-points.html); for more on polling biases and other statistical in sights see Nate Silver, The Signal and the Noise: Why So Many Predictions Fail- But Some Don’t (Penguin, .(2012

5. لتسلسل الجيني لستيف جوبز - والتر إيزاكسون وستيف جوبز
Simon andSchuster, 2011), pp. 550-)
.55l

6. توقعات نماذج انتشار الإنفلونزا على مستوى المدن - دوجاس وآخرون،
”.Google FluTrends“

7. رأي إتزيوني عن البيانات المؤقتة - مقابلة مع كوكيير، October 2011.

اقتباسات جون كونز - جوناثان روزنتال،
Special Report:“
International Banking” The Economist, May 19,2012,pp. 7-
.8

التلاعب بمباريات السومو - مارك دوجان وستيف دي ليفيت،
Winning“
Isn’t Everything: Corruption in Sumo Wrestling,” American Economic Review 92 (2002), pp. 1594-1605
.((http://pricetheory.uchicago.edu/levittjPapers/DugganLevitt2002.pdf

8. الأحد عشر مليون شعاع ضوئي من ليترو - من موقع شركة ليترو الإلكتروني
.(http://www. lytro.com)

9. استبدال أسلوب العينات في العلوم الاجتماعية - مايك سافاج وروجر بيروس،
The“
Coming Crisis of Empirical Sociology,” Sociology 41 (2007), PP- 885-
.899

10. عن تحليل البيانات الشاملة من شركات شبكات المحمول - جاي بي أونيلا وآخرون،
Structure“
and Tie Strengths in Mobile Communication Net works,” Proceedings of the National Academy of Sciences of the United States of America (PNAS) 104

(May 2007), pp. 7332-36 (http://nd.edu/-dddas/papers/pNAS0610245104vl.pdf).

3. فوضى البيانات

1. كروسبي - ألفريد دبليو كروسبي، The Measure of Reality: Quantification andWestern Society, 1250-1600 (Cambridge University Press, 1997).

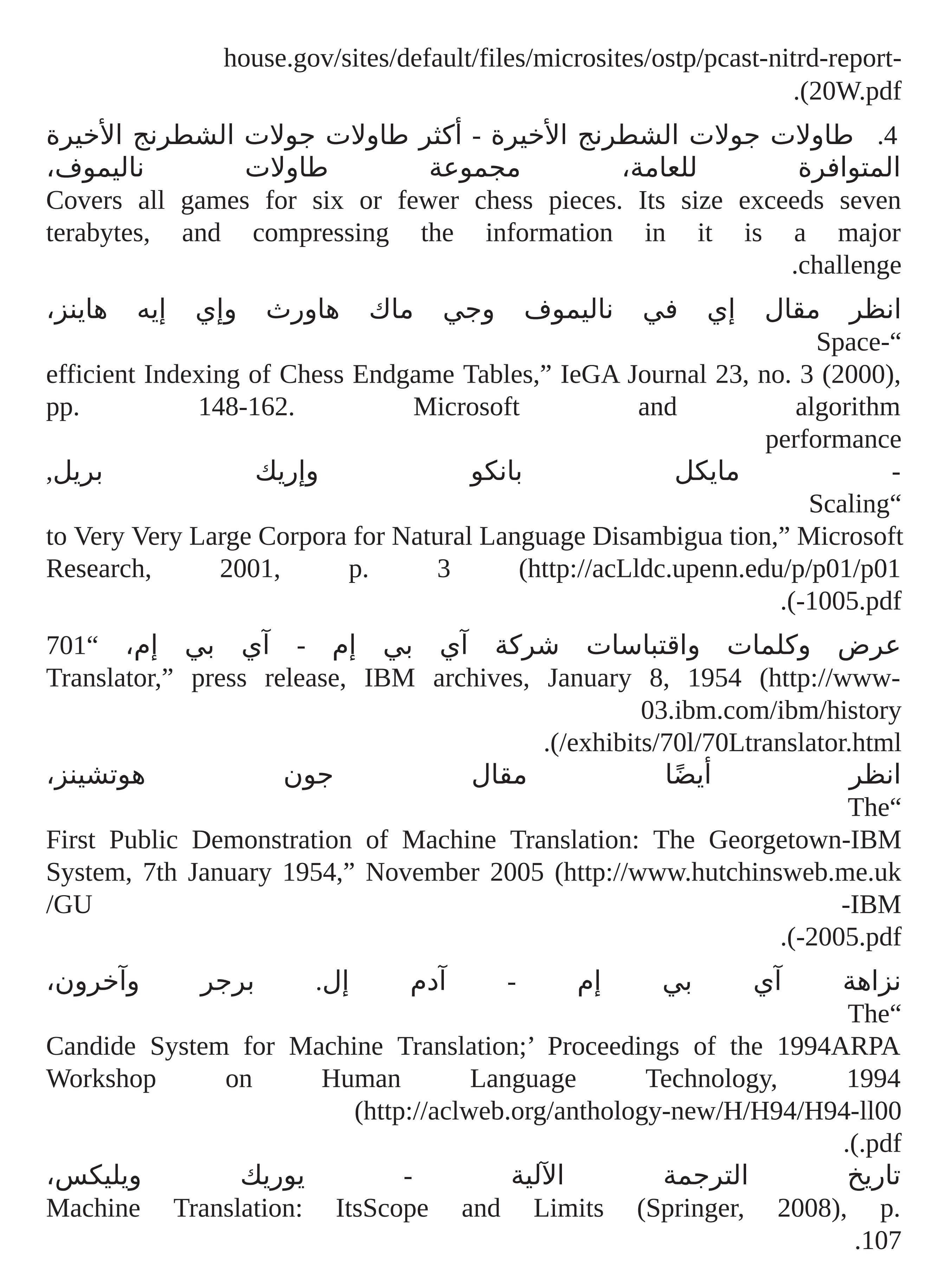
عن اقتباسات كلفين وبايكون ــــ These aphorisms are widely attributed to both men, though the actual expression in their written works is slightly different. In Kelvin, it's part of a longer quotation on measurement, from his lecture "Electrical Units of Measurement" (1883). For Bacon, it's considered to be a loose translation from Latin, in Meditationes Sacrae (1597)).

2. الكثير من الطرق للإشارة لشركة آي بي إم - دي جاي باتيل، "Data Jujitsu: The Art of Turning Data into Product," O'Reilly Media, July 2012 (http://oreillynet.com/oreilly/data/radarreports/data-jujitsu.csp?cmp=tw-strata-books-data-products).

3. 30 ألف تعامل على موقع نيس كل ثانية - كولن كلارك، "Improving Speed and Transparency of Market Data," NYSE EURONEXT blog post, January 9, 2011 (http://exchanges.nyx.com/cclark/improving-speed-and -transparency-market-data).

فكرة أن "2+2=3.9" ــــ بريان هوبكنز وبوريس إيفلسون، "Expand Your Digital Horizon with Big Data," Forrester, September 30,2011.

تحسينات الخوارزميات- مجلس الرئيس الاستشاري للعلوم والتكنولوجيا، "Report to the President and Congress, Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology," December 2010, p. 71 (http://www.white



house.gov/sites/default/files/microsites/ostp/pcast-nitrd-report-
.(20W.pdf

4. طاولات جولات الشطرنج الأخيرة - أكثر طاولات جولات الشطرنج الأخيرة المتوافرة للعامة، مجموعة طاولات ناليموف،
Covers all games for six or fewer chess pieces. Its size exceeds seven terabytes, and compressing the information in it is a major .challenge

انظر مقال إي في ناليموف وجي ماك هاورث وإي إيه هاينز، "Space-
efficient Indexing of Chess Endgame Tables," IeGA Journal 23, no. 3 (2000), pp. 148-162. Microsoft and algorithm performance
- مايكل بانكو وإريك بريل,
"Scaling
to Very Very Large Corpora for Natural Language Disambigua tion," Microsoft Research, 2001, p. 3 (http://acLldc.upenn.edu/p/p01/p01
.(-1005.pdf

عرض وكلمات واقتباسات شركة آي بي إم - آي بي إم، "701
Translator," press release, IBM archives, January 8, 1954 (http://www-03.ibm.com/ibm/history
.(/exhibits/70l/70Ltranslator.html
انظر أيضًا مقال جون هوتشينز،
"The
First Public Demonstration of Machine Translation: The Georgetown-IBM System, 7th January 1954," November 2005 (http://www.hutchinsweb.me.uk /GU -IBM
.(-2005.pdf

نزاهة آي بي إم - آدم إل. برجر وآخرون،
"The
Candide System for Machine Translation;' Proceedings of the 1994ARPA Workshop on Human Language Technology, 1994 (http://aclweb.org/anthology-new/H/H94/H94-ll00
.(.pdf
تاريخ الترجمة الآلية - يوريك ويليكس،
Machine Translation: ItsScope and Limits (Springer, 2008), p. .107

5. ملايين النصوص من كانديد مقابل مليارات النصوص من جوجل - مقابلة أوش الصحفية مع كوكيير، December 2009.

جسد جوجل المكون من 95 مليار عبارة - ألكس فرانز وثورستن برانتس، "All Our N-gram are Belong to You," Google blog post, August 3, 2006 (http://googleresearch.blogspot.co.uk/2006/08/all-our-n-gram-are-belong-to-you.html).

6. جسد براون وكلمات جوجل التي عددها 1 تريليون كلمة - هلفي ونورفيج وبيريبرا، "The Unreasonable Effectiveness of Data." اقتباس من البحث الذي شارك نورفيج في تأليفه - السابق.

7. تآكل أنابيب بي بي والبيئة اللاسلكية المعادية - جاكلين كلارابوت، "Operations Making Sense of Corrosion" BP Magazine, issue 2 (2011) (http://www.bp.com/liveassets/bp_internet/globalbp/globalbp_uLenglish/re). أتت المعلومات عن صعوبة قراءة البيانات اللاسلكية من مقال كوكيير، "Data, Data, Everywhere." The system is obviously not infallible: a fire at the BP Cherry Point refinery in February2012 was blamed on a corroded pipe.

8. مشروع المليار سعر - من مقابلة المؤسسين المشاركين مع كوكيير، October 2012. وكذلك جايمس سورويكي، "A Billion Prices Now," The New Yorker, May 30, 2011; data and details can be found on the project's web site (http://bpp.mit.edu/); أني لاوري، "Economists' Programs Are Beating U.S. at Tracking Inflation," Washington Post, December 25, 2010 (http://www.washingtonpost.com/wp-dyn/content/article/2010/12/25/AR2010122502600.html).

9. عن إحصائيات الأسعار كجزء من الإحصائيات القومية - "Official Statistics: Don't Lie to Me, Argentina," The Economist, February 25, 2012

.((http://www.economist.com/node/21548242
عدد الصور على موقع فليكر- من موقع فليكر الإلكتروني http://www.flickcom). On the challenge to categorize) information - انظر مقال دافيد فاينبرجر،
Everything Is Miscellaneous: The Power of the New Digital Disorder (Times, .(2007

10. 10 بات هيلاند - بات هيلاند، If“
You Have Too Much Data Then ‘Good Enough’ Is Good Enough,” Communications of the ACM, June 2011, pp. 40, 41. There is a vigorous debate within the database community about the models and concepts best able to meet the needs of big data. Helland represents the camp arguing for a radical break with tools used in the .past
مايكل ريس من مايكروسوفت، in “Scalable SQL;’ Communications of the ACM, June 2011, p. 48, argues that much-adapted versions of existing tools will work .fine

11. استخدام بطاقات هادوب الائتمانية - كوكيير، Data, data“ .”everywhere

12. 5 % فقط من المعلومات قائمة على البيانات-أبيهيشك ميتها، Big“
Data: Powering the Next Industrial Revolution,” Tableau Software White Paper, 2011 (http://www.tableausoftware.com/learn/whitepapers /big-data-.(revolution

4. العلاقات التبادلية

1. قصة ليندن و"صوت أمازون"- مقابلة ليندن الصحفية مع كوكييرMarch 2012.

وول ستريت جورنال تنتقد أمازون - كما ذُكر في مقال جايمس ماركوس، Amazonia: Five Years at the Epicenter of the Dot.Com Juggernaut (New Press, 2004), p. .128.51
اقتباس ماركوس - ماركوس، Amazonia, p. 199.

2. تشكل التوصيات ثلث الدخل الإجمالي لأمازون - لم يتم تأكيد هذا الرقم

بشكل رسمي من قبل الشركة ولكنه نُشر في الكثير من التقارير التحليلية والمقالات في وسائل الإعلام من بينها,
"Building with Big Data: The Data Revolution Is Changing the Landscape of Business," The Economist, May 26, 2011 (http://www.economist.com/node/18741392/).
كما تمت الاستعانة بهذه الأرقام في مقابلتين صحفيتين لكوكيير مع اثنين من المديرين التنفيذيين السابقين للشركة. معلومات أسعار نتفليكس - زافير أماترياين وجاستن بيسيليكو،
"Netflix Recommendations: Beyond the 5 stars (Part 1),"Netflix blog, April 6, 2012.

3. "خداع العشوائية" - نصيم نيكولاس طالب،
Fooled by Randomness (Random House, 2008); للمزيد، انظر كتاب نصيم نيكولاس طالب،
The Black Swan: The Impact of the Highly Improbable (2nd ed., Random House, 2010).

4. وولمارت وبوب تارت- كونستانس إل. هايز،
"What Wal-Mart Knows About Customers' Habits" New York Times, November 14, 2004 (http://www.nytimes.com/2004/ll/14/business/yourmoney/14wal.html).

5. أمثلة عن نماذج توقعات فيكو وإكسبيريان وإكويفاكس - سكوت ثورم،
"Next Frontier in Credit Scores: Predicting Personal Behavior" Wall Street Journal, October 27, 2011
line.wsj.com/article/SBI0001424052970203687504576655182086300912.html

6. نماذج توقعات أفيفا - ليزلي سيزم ومارك مارمونت،
"Insurers Test Data Profiles to Identify Risky Clients," Wall Street Journal, November19, 2010
line.wsj.com/article/SBI0001424052748704648604575620750998072986.html
انظر أيضًا مقال ليزلي سيزم ومارك مارمونت،
"In

side Deloitte’s Life-Insurance Assessment Technology;’ Wall Street Jour nal, November 19,2010
line.wsj.com/article/SBI0001424052748704104104575622531084755588.html
انظر أيضًا مقال هوارد ميلز، Analytics:“
Turning Data into Dollars,” Forward Focus, December 2011 (http://www.deloitte.com/assets/Dcom-States/Local%20Assets/Documents/FSI/US_FSLForward%20Focus_Analytics_Turning%20data%20
.(into%20dollars_1207ll.pdf

أمثلة عن شركة تارجت والمراهقات الحوامل - تشارلز دوهيج، How“
Compa nies Learn Your Secrets” New York Times, February 16, 2012 (http://www .nytimes.com/20l2/02/19/magazine/shopping-habits.html). The article is adapted from Duhigg’s book The Power of Habit: Why We Do What We Do in Life and Business (Random House, ;(20i2
أشارت شركة تارجت إلى عدم دقة معلومات وسائل الإعلام عن أنشطتها ولكنها امتنعت عن ذكر هذه المعلومات غير الدقيقة. عند سؤاله عن موضوع الكتاب، أجاب المتحدث الرسمي عن شركة تارجت: “الهدف هو استخدام بيانات العملاء من أجل تحسين علاقتهم بالشركة. يرغب عملاؤنا في الحصول على قيم كبيرة وعروض تتعلق بما يريدون وتجارب رائعة مع الشركة. إننا مثل الكثير من الشركات نستخدم أدوات البحث التي تساعدنا على فهم التوجهات التسوقية للعملاء ومفضلاتهم حتى نتمكن من تقديم عروض مناسبة لهم. إننا نتحمل مسئولية حماية ثقة عملائنا بنا على محمل الجد. من بين طرق تحقيق هذا الأمر أننا نمتلك سياسة خصوصية شاملة والتي ننشرها على موقع الشركة وأننا ندرب فريق عملنا بشكل متواصل على كيفية الحفاظ على معلومات عملائنا”.

7. عمل شركة يو بي إس التحليلي - مقابلة كوكيير مع جاك ليفيس, 2012.

8. الأطفال المبتسرون - بناءً على المقابلات الصحفية مع ماكجريجور في الفترة ما بين عامي 2010 و2012. انظر أيضًا مقال كارولين ماكجريجور وكريستينا كاتلي وأندرو جايمس وجايمس بادبري، Next“
Generation Neonatal Health Informatics with Artemis,” in Euro pean Federation for Medical Informatics, User Centred Networked Health Care, ed. A. Moen et al. (lOS Press, 2011), p. 117.Some material comes from Cukier, “Data, Data,

".Everywhere

9. عن العلاقة بين السعادة والدخل - آر إنجلهارت وإتش. دي. كلينجمان، Genes, Culture and Happiness (MIT Press, .(2000

10. عن الحصبة والتكاليف الصحية، وعن أدوات تحليل العلاقات التبادلية غير الخطية - دافيد ريشيف وآخرون، "Detecting Novel Associations in Large Data Sets," Science 334 (2011), pp. 1518-.24

11. كاهنمان - دانييل كاهنمان، Thinking, Fast and Slow (Farrar, Straus and Giroux, 2011), pp. 74-.75

12. 12 باستير - للقراء الذين يرغبون في معرفة المزيد عن تأثير باستير على إدراكنا للأمور، ننصح بقراءة كتاب برونو لا تور، The Pasteurization of France (Harvard University Press, .(1993 مخاطر الإصابة بداء الكلب - ميلان يدي كينزيو وآن ماكارثي، "Rabies Risk Among Travellers" CMAJ 178,no. 5 (2008), p. .567

13. نادرًا ما يمكن إثبات السببية - طورت عالمة الحاسب الحاصلة على جائزة تورينج "جاديا بيرل"طريقة لعرض الديناميكية السببية بشكل رسمي، رغم عدم وجود أي إثبات رسمي، فإنها تقدم أسلوبًا واقعيًّا لتحليل الصلات المحتملة سببيًّا، انظر كتاب جوديا بيرل، Causality: Models, Rea soning and Inference (Cambridge University Press, .(2009

14. مثال السيارة البرتقالية - كوينتن هاردي. "Bizarre Insights from Big Data" nytimes.com, March 28, 2012 (http://bits.blogs.nytimes.com/2012/03/28 /bizarre-insights-from-big-;(/data
وكاجل، "Momchil Georgiev Shares His Chromatic Insight from Don't Get Kicked," blog posting, February 2, 2012 (http://blog.kaggle.com/2012/02/02/momchil-georgiev-



shares-his-chromatic-insight-from -dont -get-
.(/kicked

15. تقييم أغطية فتحات الصرف وعدد الانفجارات وارتفاع الانفجار - راتشيل إهرنبرج،
Predicting“
the Next Deadly Manhole Explo sion,” Wired, July 7, 2010 (http://www.wired.com/wiredscience/201O/07/manhole-
.(explosions

عمل كون إديسون مع علماء الإحصاء من جامعة كولومبيا - تم شرح هذه الحالة للمشاركين بواسطة سينثيا أودين وآخرون،
21st-“
Century Data Miners Meet 19th-Century Electrical Cables” Computer, June 2011, pp. 103-
.105

المواصفات التقنية للعمل متوافرة في المقالات الأكاديمية لرودين وزملائها على مواقعهم الإلكترونية، وخاصة سينثيا رودين وآخرين,
Machine“
Learning for the New York City Power Grid” IEEE Transactions on Pattern Analysis and Machine Intelligence 34, no. 2 (2012), pp. 328-345
.((http://hdl.handle.net/1721.1/68634

16. فوضى مصطلح "صندوق الخدمات"- حصلنا على هذه القائمة من موقع رودين وآخرين.,
21st-“
Century Data Miners Meet 19th-Century Electrical
”.Cables الاقتباس عن رودين
- من مقابلة صحفية مع كوكيير، March 2012.

17. وجهات نظر أندرسون - كريس أندرسون،
The“
End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” Wired, June 2008 (http://www.wired.com/science/discoveries/magazine/16-
.(/07/pb_theory

18. تباطؤ أندرسون - الإذاعة القومية العامة،
Search“
and Destroy:’ July18, 2008 (http://www.onthemedia.org/2008/jul/18/search-and-destroy
.(//transcript

.19 عن الخيارات المؤثرة على تحليلاتنا - دانا بويد وكايت كراوفورد. Six“
Provocations for Big Data,” paper presented at Oxford Internet Institute’s “A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society” September 21, 2011 .((http://ssrn.com/abstract=1926431

5. التحول إلى البيانات

1. تم جمع المعلومات عن حياة ماوري من الكثير من الأعمال التي تتحدث عنه أو التي ألفها بنفسه، ومن بينها كتاب تشستر جي هيرن،
Tracks in the Sea: Matthew Fontaine Maury and the Mapping of the Oceans (International Marine/McGraw-Hill, ;(2002
جانيس بيتي،
Seeker of Seaways: A Life of Matthew Fontaine Maury, Pioneer Oceanographer (pantheon Books, ;(1966
تشارلز لي لويس،
Matthew Fon taine Maury: The Pathfinder of the Seas (U.S. Naval Institute, 1927) (http:// ;(archive.org/details/matthewfontainemOOlewi
وماتيو فونتاين ماوري، The Physical .(Geography of the Sea (Harper, 1855

2. الاقتباسات عن ماوري - من كتاب ماوري،
Physical Geography of the Sea, “Introduc- tion;’ pp. xii, .vi

3. بيانات مقاعد السيارات - نيكي،
Car Seat of Near“ Future IDs Driver’s Backside” December .14,2011

4. قياس العالم - أغلب فكر المؤلف يعتمد على أن تاريخ سيطرة البيانات قد قدمه كروزبي، The Measure of Reality.

5. لم يتعلم الأوروبيون الحساب على المعداد - المرجع السابق,. 112. الحساب بطريقة أسرع من خلال استخدام الأرقام العربية - ألكسندر ماوري،
Reason and Society in the Middle Ages (Oxford University Press, 1978), p. .166

6. عدد الكتب المنشورة ودراسات جامعة هارفارد في مشروع جوجل للمسح الضوئي للكتب - جون بابتيست مايكل وآخرون، "Quantitative Analysis of Cul ture Using Millions of Digitized Books" Science 331(January 14,2011), pp. 176-182 ((http://www.sciencemag.org/content/331/6014/176.abstract).

للحصول على محاضرة مصورة عن الدراسة، انظر كتاب إيريز ليبرمان إيدن وجون بابتيست مايكل، "What We Learned from 5 Million Books" TEDx, Cambridge, MA, 2011 (http://www.ted.com/talks/whaLwe_learnecLfroI1L5_million.books.html).

7. عن النماذج اللاسلكية للسيارات والتأمين - انظر كوكيير، "Data, DataEverywhere."

جاك ليفيس من شركة يو بي إس - مقابلة مع جاك كوكيير، April 2012.

تأثير البيانات على توفير تكاليف يو بي إس - معهد أبحاث التشغيل وعلوم الإدارة (INFORMS) "UPS Wins Gartner BI Excellence Award," 2011 (http://www.informs.org/Announcements/UPS-wins-Gartner-BI-Excellence-Award).

8. بحث بنتلاند - روبرت لي هوتز، "The Really Smart Phone," Wall Street Journal, April 22, 2011 (http://online.wsj.com/article/SBlO0014240527487045476045762632616798488 14.html).

9. دراسة إيجل عن الأحياء الفقيرة - ناثان إيجل، "Big Data, Global Development, and Complex Systems," Santa Fe Institute, May 5, 2010 (http://www.youtube.com/watch?v=yaivtqlu7iM). Also, interview with Cukier, October 2012.

10. بيانات فيس بوك - من محاكمات الاكتتاب العام لشركة فيس بوك, 2012.

بيانات تويتر - أليكسيا تسوتسيتس، "Twitter Is at250 Million Tweets per Day, iOS5 Integration Made Signups Increase 3x,"

TechCrunch, October 172011, http:// techcrunch.com/20ll/10 /17/twitter- is-at-250-million-tweets-per-
./day

المحافظ الوقائية باستخدام تويتر - كينيث كوكيير، Tracking"
Social Media: The Mood of the Market" Economist.com, June 28, 2012 (http://www .economist.com/blogs/graphicdetail/2012/06/tracking-social-
.(media

11. تويتر وتوقع أرباح أفلام هوليوود - ستارام أسور وبرناردو إيه هوبرمان، Predicting"
the Future with Social Media" Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 492-499; online at http://www.hpl.hp.com/research/scl/papers/socialmedia/socialmedia
..pdf

تويتر والمزاج العالمي - سكوت إيه. جولدر ومايكل دبليو. ماسي، Diur"
nal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Di verse Cultures," Science 333 (September 30, 2011), pp. 1878-
.81

تويتر والتطعيم ضد الإنفلونزا - مارسل سالاذيه وشاشانك كاندلوال، Assess"
ingVaccination Sentiments with Online Social Media: Implications for In fectious Disease Dynamics and Control," PLoS Computational Biology, Oc tober
.2011

12. اختراع آي بي إم للأرضيات الذكية - ليندا ماي دو وترافيس إم. جريجسبي وباميلا آن نيسبيت وليزا آن سيكات. Securing"
premises using surfaced-based computing technology" U.S. Patent number: 8138882. Issue date: March
.12

الحركات الذاتية المقاســة Counting"
Every Moment" The Econo mist, March 3,
.2012

سماعات الأذن من شركة آبل والقياسات الحيوية- جيسي لي درورجاسكر وأنتوني فادل ودونالد جاي نوفوتني ونيكولاس آر. كالايجيان، "Integrated Sensors for Tracking Performance Metrics," U.S. Patent Application 20090287067. Assignee: Apple. Application Date: 2009-07-23. Publication Date: 2009-11-19.

قياسات ديراوي الحيوية، "Your Walk Is Your PIN-Code," press release, February 21,2011 (http://biometrics.derawi.com/?p=175).

معلومات المصطلح الذكي انظر مشروع المصطلح الذكي على صفحة مركز http://eosl.gtri.gatech.edu/Capabilitie/LandmarcResearchCenter/LandmarcProject exchange.

باحثو كيوتو وعدادات السرعة ثلاثية المحاور - فريق عمل آي ميديكال آبس، "Gait Analysis Accuracy: Android App Comparable to Standard Accelerometer Methodology" mHealth, March 23, 2012.

13. دفعة الصحف للحالة القومية - بندكت أندرسون، ImaginedCommunities: Reflections on the Origin and Spread of Nationalism (Verso, 2006).

يقول علماء الفيزياء إن المعلومات هي كل شيء - هانز كريستيان فون باير، Information: The New Language of Science (Harvard University Press, 2005).

6. القيمة

1. قصة لويس فون أهن - بناءً على المقابلة الصحفية التي أجراها كوكيير مع فون أهن عام 2010، انظر أيضًا مقال كليف تومبسون، "For Certain Tasks, the Cortex Still Beats the CPU" Wired, June 25, 2007 (http://www.wired.com/techbiz/it/mag azine/15-07/fLhumancomp?currentPage=all);

جيسي سكانلون،

Luis“
von Ahn: The Pioneer of ‘Human Computation;” Businessweek, November 3, 2008 (http://www.businessweek.com/stories/2008-11-03/luis-von-ahn-the-pioneer-of-human-computation-businessweek-business-news-stock-market-and-financial-advice). His technical description of reCaptchas is at Luis von Ahn et aI., “reCAPTCHA: Human-Based Character Rec ognition via Web Security Measures” Science 321 (September 12,2008), pp. 1465-68 .((http://www.sciencemag.org/content/321/5895/1465.abstract

2. مصنع سميث للدبابيس ــــ آدم سميث، The Wealth of Nations (reprint, Bantam Classics, 2003), book I, chapter one. (A free electronic version is at http ://www2.hn.psu.edu/faculty/jmanis/adam-smith/Wealth-
.(Nations.pdf

3. التخزين - فيكتور ماير شونبرجر، Delete: The Virtue of Forgetting in the Digital Age (Princeton University Press, 2011), p.
.63

4. عن استهلاك السيارات الكهربية للطاقة ــــ IBM, “IBM, Honda, and PG&E Enable Smarter Charging for Electric Vehicles” press release, April 12,2012 (http:// www-
.(03.ibm.com/press/us/en/pressrelease/37398.wss
انظر أيضًا كتاب كلاي لوثي، Guest“
Perspective: IBM Working with PG&E to Maximize the EV Po tential” PGE Currents Magazine, April 13, 2012 (http://www.pgecurrents .com/2012/04/13/ibm-working-with-pge-to-maximize-the-ev-
.(potential

5. بيانات أمازون وإية أو إل - مقابلة كوكيير مع أندرياس ويجنيد، 2010 and 2012. نوانس للبرمجيات وجوجل - كوكيير،“Data, Data Everywhere” .

6. الشركة اللوجيستية - براد براون ومايكل تشوي وجايمس مانييكا، AreYou“
Ready for the Era of ‘Big Data’?” McKinsey Quarterly, October 2011, p. .10

7. تسييل تليفونيكا لمعلومات الهواتف المحمولة ــــ Telefonica“

Hopes ‘Big Data’ Arm will Revive Fortunes” BBC Online, October 9, 2012. (http://www.bbc .co.uk/news/technology-19882647).

دراسة الجمعية الدنماركية للسرطان - باتريزيا فري وآخرون، “Use of Mobile phones and Risk of Brain Tumours: Update of Danish Cohort Study” BMJ 343 (2011) (http://www.bmj.com/content/343/bmj.d6387). and interview with Cukier, October 2012.

8. تسجيلات شركة جوجل لاستعراض الشوارع والسيارات ذاتية القيادة - بيتر كيروان، “This Car Drives Itself ” Wired UK, January 2012 (http://www.wired.co.uk /magazine/archive/2012/01/features/this-car-drives-itself?page=all).

9. برنامج جوجل للتدقيق الإملائي والاقتباسات - مقابلة صحفية مع كوكيير في المقر الرئيسي لجوجل بماونتن فيو بكالفورنيا، December 2009; some material also ap peared in Cukier, “Data, Data Everywhere.”

10. أفكار هامرباخر - مقابلة مع كوكيير، October 2012. Barnes & Noble e-book data-Alexandra Alter, “Your E-Book Is Reading You” Wall Street Journal, June 29, 2012 (http://online.wsj.com/article/SB10001424052702304870304577490950051438304.html).

11. بيانات ودراسة أنردو إن جي - مقابلة مع كوكيير، June 2012.

12. سياسة أوباما للحكومة المنفتحة- باراك أوباما، “White Presidential memo randum” House, January 21, 2009.

13. عن قيمة بيانات فيس بوك ــ بالنسبة للاستعراض الرائع عن الاختلاف بين قيمة الاكتتاب العام في السوق والقيمة الدفترية له، انظر مقال دو جلاني، “To Facebook You’re Worth $80.95,” Wall Street Journal, May 3, 2012 (http://blogs.wsj.com/cio/2012/05/03/to-facebook-youre-worth-80-95/).

لتقييم البنود الخفية لشركة فيس بوك، قدر لاني استقرائيًّا من نمو الشركة حوالي 2.1 تريليون من قصاصات المحتوى. في مقاله في صحيفة وول ستريت جورنال، قدر لاني قيمة كل من هذه القصاصات بثلاثة سنتات لكل منها، وكان يُستخدم تقييم سوقي مقدر سابق لفيس بوك بقيمة 75 مليار دولار. في النهاية، وصلت قيمتها إلى ما يزيد على 100 مليار دولار، أو 5 سنتات لكل قصاصة محتوى، كما سمحنا لأنفسنا بأن نقدر بناءً على حساباته. الفرق في القيمة بين الأصول المادية والملموسة، انظر مقال ستيف إم. سامك، "Pre

pared Testimony: Hearing on Adapting a 1930's Financial Reporting Model to the 21st Century," U.S. Senate Committee on Banking, Housing and Ur ban Affairs, Subcommittee on Securities, July 19,2000.

قيمة الأصول غير الملموسة- روبرت إس. كابلان ودافيد بي نورتون، Strategy Maps: Converting Intangible Assets into Tangible Outcomes (Harvard Business Review Press, 2004), pp. 4-5.

14. الاقتباس عن تيم أورايلي - مقابلة مع كوكيير، February 2011.

7. ما تتضمنه البيانات

1. معلومات عن موقع ديسيد دوت كوم ــ رسائل بريد إلكتروني بين كوكيير وإتزيوني، May 2012.

2. تقرير ماكينزي ــ جايمس مانييكا وآخرين، "Big

Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute, May 2011 n/insights/mgi/research/technology_and_innovation/big_data_the_nexLfrontier.for.fnnovation), p. 10.

الاقتباس عن هال فاريان - مقابلة مع كوكيير، December 2009.

3. الاقتباس عن كارل دي ماركن - مراسلات إلكترونية مع كوكيير، May 2012.

4. عن مستشاري ماستر كارد - مقابلة كوكيير الصحفية مع جاري كيرنز، رئيس مستشاري ماستر كارد، At The Economist's "The Ideas Economy: Infor mation" conference, Santa Clara, California, June 8, 2011.

5. شركة أسينتور ومدينة سانت لويس بولاية ميسوري - مقابلة كوكيير مع بعض موظفي المدينة، February 2007. نظام أمالجا الموحد للمعلومات، "Microsoft
Expands Pres ence- in Healthcare IT Industry with Acquisition of Health Intelligence Software Azyxxi," Microsoft press release, July 26, 2006 (http://www .microsoft.com/en-us/news/press/2006/jul06/07-26azyxxiacquisitionpr.aspx). The Amalga service is now a part of Microsoft's joint venture with General Electric called .Caradigm

6. بردافورد كروس - مقابلات مع كوكيير، March-October .2012

7. أمازون والترشيح التعاوني - دعاوى الاكتتاب العام، May 1997 (http://www.sec.gov/Archives/edgar/data/l018724/0000891020-97-.000868

8. المعالجات الدقيقة للسيارات - نيك فاليري، "Tech.View:
Cars and Software Bugs," Economist.com, May 16, 2010 ww.economist.com/blogs/babbage/2010/05/techview_cars_ancLsoftware_bugs

أطلق ماوري على السفن "المراصد العائمة" - ماوري، .The Physical Geography of the Sea

9. إنريكس - مقابلة كوكيير مع المديرين التنفيذيين، May and .September, 2012

10. 10 عن معهد تكاليف الرعاية الصحية - سارة كليف، "A
Database That Could Revolutionize Health Care," Washington Post, May 21, 2012. Decide.com's data-usage agreement - Cukier email exchange with Etzioni, May .2012

صفقة جوجل وآي تي إيه- كلير كاين ميللر، "U.S.
Clears Google Acquisition of Travel Software," New York Times, April 8, 20ll (http://www.nytimes .com/2011/04/09 /technology .(/09google.htmILr=0

11. إنريكس وإيه بي إس - مقابلة كوكيير والمديرين التنفيذيين لشركة إنريكس،

.May 2012

12. قصة القوارض والاقتباس عن لين كينيدي - مقابلة مع كوكيير, May 2012.
حوار من فيلم موني بال من تأليف بينيت ميللر، Columbia Pictures, 2011.

13. زيادة بيانات ماكجريجور إلى ما يتخطى عقدًا كاملاً من أعوام المرضى - مقابلة مع كوكيير، May 2012.

اقتباس عن جولد بلووم - مقابلة مع كوكيير، March 2012.

14. عن إيراد أفلام هوليوود ومبيعات الألعاب - عن الأفلام انظر مقال بروكس بارنز،
A“
Year of Disappointment at the Movie Box Office,” New York Times, December 25, 20ll (http://www.nytimes.com/2011/12/26/business/media/a-year-of-disappointment-for-hollywood.html).
عن ألعاب الفيديو انظر
Factbox:“
A Look at the $65 billion Video Games Indus try,” Reuters, June 6, 2011 (http://uk.reuters.com/articie/2011/06/06 /us-videogames-factbox-idUKTRE75552I20110606).

تحليل بيانات شركة زينجا - نيك وينجفيلد،
Virtual“
Products, Real profits: Players Spend on Zynga’s Games, but Quality Turns Some Off,” Wall Street Journal, September 9, 2011 (http://online.wsj.com/articie/SBI0001424053111904823804576502442835413446.html).

15. الاقتباس عن كين رودين - من مقابلة مع رودين أجراها نيكو وايش واستعان بها إريك شلي ويورج راينبولدت ونيكو وايش في كتابهم،
Simply Seven: Seven Ways to Create a Sustainable Internet Business (Palgrave Macmillan, 2011). p. 7. Auden quotation - W. H. Auden, “For the Time Being,” 1944.

الاقتباس عن توماس دافنبورت - مقابلة أجراها كوكيير مع دافنبورت، December 2009.

موقع نمبرز دوت كوم - مقابلة أجراها كوكيير مع بروس ناش، October 2011 and July 2012.

16. دراسة برينجولفسون- إريك برينجولفسون ولورين هيت وهيكيونج كيم، "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" working paper, April 2011 (http://papers.ssrn.com/soI3/papers.cfm?abstracLid=1819486).

17. عن رولزرويس؛ انظر "Rolls-Royce: Britain's Lonely High- Flier" The Econ- omist, January 8, 2009 (http://www.economist.com/node/12887368). Fig- ures updated from press office, November 2012.

إريك برينجولفسون وأندرو ماكافي ومايكل سوريل وفينج زهو، "Scale Without Mass: Business Process Replication and Industry Dynamics," Harvard Business School working paper, September 2006 (http://www.hbs.edu/research/pdf/07-0l6.pdf also http://hbswk.hbs.edu/item/5532.tml).

18. عن التوجه نحو مالكي البيانات الكبار - انظر أيضًا كتاب يانيس باكوس وإريك برينجولفسون، "Bundling Information Goods: Pricing, Prof its, and Efficiency" Management Science 45 (December 1999), pp. 1613-30.

19. فيليب إيفانز - مقابلة مع المؤلفين، 2011 and 2012.

8. المخاطر

1. عن البوليس السري الألماني ــ لسوء الحظ، كانت أغلب المعلومات باللغة الألمانية، باستثناء وحيد هو كتاب كريستي ماكراكيس، Seduced by Secrets: Inside the Stasi's Spy- Tech World (Cambridge University Press, 2008); a very per sonal story is shared in Timothy Garton Ash, The File (Atlantic Books, 2008). We also recommend the Academy Award-winning movie The Lives of Others, directed by Florian Henckel von Donnersmark, Buena Vista/ Sony Pictures, 2006.

كاميرات المراقبة بالقرب من منزل أورويل، George“
Orwell, Big Brother Is Watching Your House,” The Evening Standard, March 31, 2007 (http:// www.thisislondon.co.uk/news/george-orwell-big-brother-is-watching-your-house-
.(7086271.html

عن إكويفاكس وإكسبيريان ـــ
Daniel J. Solove, The Digital Person: Technology and Privacy in the Information Age (NYU Press, 2004), pp. 20-
.2l

2. عن عناوين منازل اليابانيين في واشنطن والتي تم تسليمها إلى السلطات الأمريكية - جاي آر مينكل،
The“
U.S. Census Bureau Gave Up Names of Japanese Americans in WW II,” Scientific American, March 30, 2007
(http://www.scientificamerican.com/article.cfm?id=confirmed-the-us-census-
.(b

عن البيانات التي استخدمها النازيون في هولندا - ويليام سيلتزر ومارجو أندرسون،
The“
Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses” Social Research 68 (2001), pp. 481-
.513

3. عن آي بي إم والهولوكوست- إدوين بلاك، IBM
.(and the Holocaust (Crown, 2003 عن كم البيانات
التي جمعتها المقاييس الذكية - انظر مقال إلياس لييك كوين،
Smart“
Metering and Privacy: Existing Law and Competing Policies; A Re port for the Colorado Public Utility Commission” Spring 2009
.((http://www.w4ar.comjDangeLoLSmarLMeters_Colorado_Report.pdf
انظر أيضًا مقال جويل إم. مارجوليس،
When“
Smart Grids Grow Smart Enough to Solve Crimes” Neustar, March 18, 2010
(http://energy.gov/sites/prod/files/gc
(prod/documents/Neustar_Comments_DataExhibitA.pdf

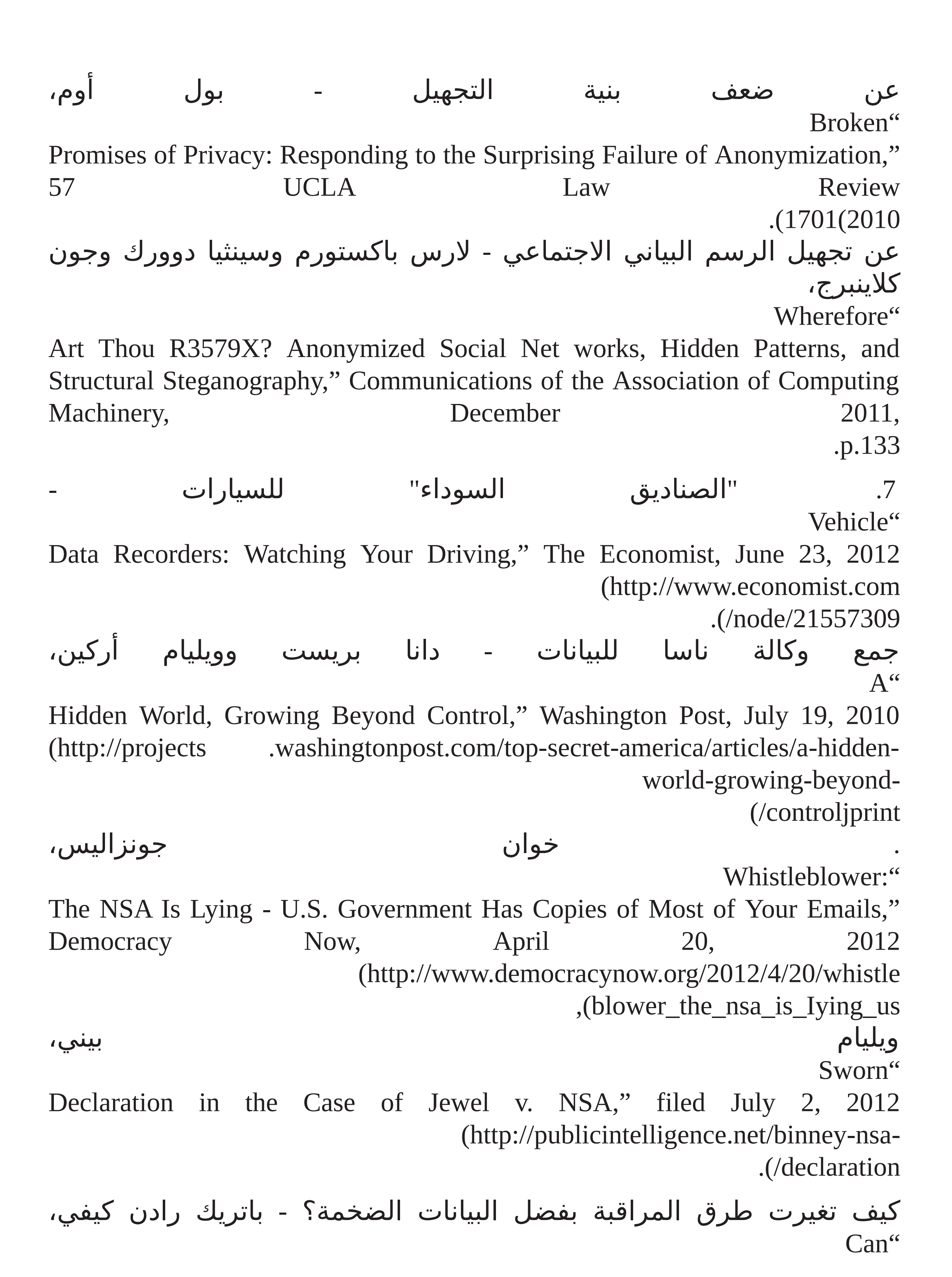
4. فريد كايت عن الملاحظة والقبول- فريد إتش. كايت, "The Failure of Fair Information Practice Principles". في كتاب جين كاي وين, Consumer Protection in the Age of the "Information Economy" (Ashgate, 2006), p. 341 et seq.

5. عن نشر شركة إيه أو إل للبيانات - مايكل باربارو وتوم زيللر جونيور, "A Face Is Exposed for AOL Searcher No. 4417749" New York Times, August 9, 2006.
انظر أيضًا مقال ماتيو كارينتشينج ومايلين مانجاليندان, "AOL Fires Technology Chief After Web-Search Data Scandal," Wall Street Journal, August 21, 2006.

6. مستخدمي نت فليكس المحددة هوياتهم - ريان سينجل، "Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims," Wired, December 17, 2009 (http://www.wired.com/threatlevel/2009 /12/netflix-privacy-lawsuit/).

عن نشر شركة نت فليكس للبيانات - أرفيند نارايانان وفيتالي شماتيكوف، "Robust De-Anonymization of Large Sparse Datasets," Proceedings of the 2008 IEEE Symposium on Security and Privacy, p. 111 et seq. (http://www.cs.utexas.edu/-shmat/shmaLoak08netflix.pdf);
أرفيند نارايانان وفيتالي شماتيكوف، "How to Break the Anonymity of the Netflix Prize Dataset," October 18, 2006, arXiv:cs/0610105 [cs.CR] (http://arxiv.org/abs/cs/0610105)).

عن تحديد هويات المستخدمين من خلال سمات ثلاث - فيليب جول، "Revisiting the Uniqueness of Simple Demographics in the US Population," Association for Computing Machinery Workshop on Privacy in Electronic Society 5 (2006), p. 77.



عن ضعف بنية التجهيل - بول أوم،
Broken“
Promises of Privacy: Responding to the Surprising Failure of Anonymization,” 57 UCLA Law Review .(1701)2010
عن تجهيل الرسم البياني الاجتماعي - لارس باكستورم وسينثيا دوورك وجون كلاينبرج،
Wherefore“
Art Thou R3579X? Anonymized Social Net works, Hidden Patterns, and Structural Steganography,” Communications of the Association of Computing Machinery, December 2011, .p.133

.7 "الصناديق السوداء" للسيارات -
Vehicle“
Data Recorders: Watching Your Driving,” The Economist, June 23, 2012 (http://www.economist.com .(/node/21557309
جمع وكالة ناسا للبيانات - دانا بريست وويليام أركين،
A“
Hidden World, Growing Beyond Control,” Washington Post, July 19, 2010 (http://projects .washingtonpost.com/top-secret-america/articles/a-hidden-world-growing-beyond- (/controljprint
. خوان جونزاليس،
Whistleblower:“
The NSA Is Lying - U.S. Government Has Copies of Most of Your Emails,” Democracy Now, April 20, 2012 (http://www.democracynow.org/2012/4/20/whistle ,(blower_the_nsa_is_Iying_us
ويليام بيني،
Sworn“
Declaration in the Case of Jewel v. NSA,” filed July 2, 2012 (http://publicintelligence.net/binney-nsa- .(/declaration

كيف تغيرت طرق المراقبة بفضل البيانات الضخمة؟ - باتريك رادن كيفي،
Can“

Network Theory Thwart Terrorists?” New York Times, March 12, 2006 .((http://www.nytimes.com/2006/03/12/magazine/312wwln_essay.html

8. حوار من فيلم مينيوريتي ريبورت, من إخراج ستيفن سبيلبرج، إنتاج مشترك بين شركتي دريم ووركس وتونتيث سينشري فوكس، 2002. تم اختصار الحوار الذي اقتبسناه قليلاً. الفيلم مقتبس من قصة قصيرة من عام 1958 من تأليف فيليب كيه. ديك، ولكن توجد بعض التغيرات الجوهرية بين النسختين. خاصة المشهد الأول للزوج الذي خانته زوجته والذي لم يكن موجودًا في الكتاب، كما أن لغز استباق الجرائم تم عرضه في فيلم سبيلبرج بصورة أكثر تشددًا عن القصة. لذا فإننا اخترنا أن نستشهد بالفيلم.

أمثلة على استخدام التوقعات في أعمال الشرطة - جايمس فلاهوس، The“ Department of Pre-Crime,” Scientific American 306 (January 2012), pp. 62-.67

9. عن تقنية استعراض الصفات المستقبلية (FAST) انظر مقال شارون واينبرجر Terrorist“ ‘Pre-crime’ Detector Field Tested in United States,” Nature, May 27, 2011 ;((http://www.nature.com/news/2011/110527/full/news.2011.323.html شارون واينبرجر “Intent to Deceive,” Nature 465 (May 2010), pp. 412-.415 عن مشكلة الإيجابيات الزائفة, انظر مقال ألكسندر فورناس“Homeland Security’s ‘Pre-Crime’ Screening will Never Work,” The Atlantic Online, April 17, 2012 (http:// www.theatlantic.com/technology/archive/2012/04/homeland-securitys-pre-crime-screening-will-never-.(/work/255971

10. عن علامات الطلبة والأقساط التأمينية - تيم كويري، Grade“ Inflation and the Good-Student Discount,” Contingencies Magazine, American Academy of Actuaries, May-June 2007 (http://www.contingencies.org .(/mayjun07/tradecraft.pdf

عن مخاطر الانطباعات العامة - برنارد إي. هاركورت، Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age (University of Chicago Press, .(2006

11. عن أعمال ريتشارد بيرك - ريتشارد بيرك، The“
Role of Race in Forecasts of Violent Crime,” Race and Social Problems 1 (2009), pp. 231-242, and email interview with Cukier, November .2012

12. عن حب ماكنامارا للبيانات - فيل روزينزويج، Robert“
S. McNamara and the Evolution of Modern Management,” Harvard Business Review, December 2010 (http://hbr.org/201O/12/robert-s-mcnamara-and-the-evolution-of-modern-
.(management/ar/pr

13. عن نجاح الأطفال البارعين خلال الحرب العالمية الثانية - جون بيرني, .(The Whiz Kids (Doubleday, 1993

عن ماكنامارا في فورد - دافيد هلبرستام، The Reckoning (William Mor row, 1986), pp. 222-.245

14. كتاب كينارد - دوجلاس كينارد، The War Managers (University Press of New England, 1977), pp. 71-.25 لقد
استفدنا في هذا القسم كثيرًا من المقابلة الإلكترونية مع دكتور كينارد، من خلال مساعده، والتي يقدم المؤلفون له بفضلها جزيل الشكر.

15. عن اقتباس "نؤمن بالله ـــ والجميع يعطوننا البيانات"- عادة ما ترجع هذه المقولة إلى دبليو. إدوارد ديمينج

عن تيد كينيدي وقائمة المنع من السفر- سارة كيهولاني جو, Sen.“
Kennedy Flagged by NO-Fly List,” Washington Post, August 20, 2004, p. A01 (http://www.washingtonpost.com/wp-dyn/articles/A17073-
.(2004Aug19.html

16. توظيف جوجل للممارسين - انظر كتاب دوجلاس إدوارد I’m Feeling Lucky: The Confessions of Google Employee Number 59 (Houghton Mifflin Harcourt, 2011), p. .9
، انظر أيضًا كتاب ستيفين ليفاي In the Plex (Simon and Schuster, 2011), pp. 140-

.141 الأمر المثير للسخرية هو أن مؤسسي شركة جوجل أرادوا تعيين ستيف جوبز مديرًا تنفيذيًّا لها despite his lack of a college degree); Levy,) .p. 80

اختبار 41 درجة من اللون الأزرق- لاورا إم. هولسون، "Putting a Bolder Face on Google," New York Times, March 1, 2009 (http://www.nytimes .com/2009 /03 .(/01/business/01marissa.html

استقالة كبير المصممين بشركة جوجل - تم اقتباس هذه المقولة (دون نقصان) من دوج بومان، "Goodbye, Google," blog post, March 20, 009 (http://stopdesign.com/archive/2009/03/20/goodbye- .(google.html

.17 اقتباس جوبز — ستيف لور، "Can Apple Find More Hits Without Its Tastemaker?" New York Times, January 18, 2011, p. B1 (http://www.ny .(times.com/20ll/01/19/technology/companies/19innovate.htm1

كتاب سكوت - جايمس سكوت، Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed (Yale University Press, .(1998

اقتباس ماكنامارا عام 1967- من خطاب ألقاه في كلية ميلسابس في مدينة جاكسون بولاية ميسيسيبي، وتم اقتباسها في .Harvard Business Review, December 2010

.18 عن دفاع ماكنامارا - روبرت إس. ماكنامارا مع بريان فانديمارك، In Retrospect: The Tragedy and Lessons of Vietnam (Random House, 1995), pp.48, .270

9. السيطرة

.1 عن مجموعة كتب جامعة كامبريدج - مارك دروجين، Anathema! Medieval Scribes and the History of Book Curses (Allanheld and

Schram, 1983), p. .37

2. عن المسئولية والخصوصية-قام مركز معلومات القادة السياسيين بمشروع ممتد على مدار سنوات عن واجهة المسئولية والخصوصية، انظر http://www.informationpolicycentre.com/accountability-based_privacy_governance/.

3. عن تواريخ صلاحية البيانات - ماير شونبرجر، Delete. "Differential privacy" - Cynthia Dwork, "A Firm Foundation for Private Data Analysis," Communications of the ACM, January 2011, pp. 86-95.

عن الفيسبوك والخصوصية إيه. تشين وإيه. كلينفلتر، "Differential Privacy as a Response to the Reidentification Threat: The Facebook Advertiser Case Study," 90 North Carolina Law Review 1417 (2012);
إيه. هايبرلن وآخرون، "Differential Privacy Under Fire," http://www.cis.upenn.edu/-ahae/papers/fuzz-sec2011.pdf.

4. الشركات المتهمة بالاحتكار - هناك عمل يجري بالفعل عن هذا الموضوع، انظر مقال بيم هايجنن وماركو إيه. هان وأدريان آر. سويتفنت، "Screening for Collusion: A Spatial Statistics Approach," Discussion Paper TI 2012-058/1, Tinbergen Institute, The Netherlands, 2012 (http://www.tinbergen.nl/discussionpapers/12058.pdf).

5. عن ممثلي الشركات الألمانية لحماية البيانات - فيكتور ماير شونبرجر، "Beyond Privacy, Beyond Rights: Towards a 'Systems' Theory of Information Governance," 98 California Law Review 1853 (2010).

6. عن قابلية التشغيل البيني- جون بلافري وأورس جاسر، Interop: The Promise and Perils of Highly Interconnected Systems (Basic Books, 2012).

10. ما سيحدث لاحقًا

1. مارك فلاورز وتحليلات مدينة نيويورك - بناءً على مقابلة مع كوكيير تمت في يوليو 2012، لوصف أفضل انظر مقال أليكس هوارد، "Predictive data analytics is saving lives and taxpayer dollars in New York City," O'Reilly Media, June 26, 2012 (http://strata.oreilly.com/2012/06/predictive-data-analytics-big-data-nyc.html).

2. وولمارت وبوتارت- هايز، "What Wal-Mart Knows About Customers' Habits!"

3. استخدام البيانات الضخمة مع الأحياء الفقيرة وتحركات اللاجئين - ناثان إيجل، "Big Data, Global Development, and Complex Systems," http://www.youtube.com/watch?v=yaivtqlu7iM.

إدراك الوقت - بندكت أندرسون، Imagined Communities (Verso, 2006)

4. "إن ما مضى هو التمهيد لما سيحدث"- ويليام شكسبير، "The Tempest," Act 2, Scene 1

5. تجربة سيرن لتخزين البيانات - مراسلات إلكترونية بين كوكيير وباحثي سيرن في نوفمبر 2012. نظام حاسب أبوللو 11 - دافيد إيه. ميندل، Digital Apollo: Human and Machine in Spaceflight (MIT Press, 2008).

المراجع

Alter, Alexandra. “Your E-Book Is Reading You.” Wall Street Journal, June 29, 2012 nline.wsj.com/article/SB10001424052702304870304577490950051438304.html

.Anderson, Benedict. Imagined Communities, New Edition. Verso, 2006

Anderson, Chris. “The End of Theory.” Wired 16, issue 7 (July 2008 .((http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Asur, Sitaram, and Bernardo A. Huberman. “Predicting the Future with Social Media.” Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 492-499. (An online version is available at (.http://www.hpl.hp.com/research/scl/papers/socialmedia/socialmedia.pdf

Ayres, Ian. Super Crunchers: Why Thinking-By-Numbers Is the New Way to .Be Smart. Bantam Dell, 2007

.Babbie, Earl. Practice of Social Research, 12th ed. 2010

Backstrom, Lars, Cynthia Dwork, and Jon Kleinberg. “Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural .Steganography.” Communications of the ACM, December 2011, pp. 133-141

Bakos, Yannis, and Erik Brynjolfsson. “Bundling Information Goods: Pricing, Profits, and Efficiency.” Management Science 45 (December 1999), pp. .1613-30

Banko, Michele, and Eric Brill. “Scaling to Very Very Large Corpora for Natural Language Disambiguation.” Microsoft Research, 2001, p. 3 .((http://acl.ldc.upenn.edu/P/P01/P01-1005.pdf

Barbaro, Michael, and Tom Zeller Jr. “A Face Is Exposed for AOL Searcher .1 No. 4417749.” New York Times, August 9, 2006 .((http://www.nytimes.com/2006/08/09/technology/09aol.html

Barnes, Brooks. “A Year of Disappointment at the Movie Box Office,” New .1 York Times, December 25, 2011 (http://www.nytimes.com/2011/12/26/business/media/a-year-of- .(disappointment-for-hollywood.html

Beaty, Janice. Seeker of Seaways: A Life of Matthew Fontaine Maury, .1 .Pioneer Oceanographer. Pantheon Books, 1966

Berger, Adam L., et al. “The Candide System for Machine Translation.” .1 Proceedings of the 1994 ARPA Workshop on Human Language Technology .((1994) (http://aclweb.org/anthology-new/R/R94/H94-1100.pdf

Berk, Richard. “The Role of Race in Forecasts of Violent Crime.” Race and .1
.Social Problems 1 (2009), pp. 231-242

.Black, Edwin. IBM and the Holocaust. Crown, 2003 .1

Boyd, danah, and Kate Crawford. “Six Provocations for Big Data.” Research .1
paper presented at Oxford Internet Institute’s “A Decade in Internet Time:
Symposium on the Dynamics of the Internet and Society,” September 21, 2011
.((http://ssrn.com/abstract=1926431

Brown, Brad, Michael Chui, and James Manyika. “Are You Ready for the .1
.Era of ‘Big Data’?” McKinsey Quarterly, October 2011, p. 10

Brynjolfsson, Erik, Andrew McAfee, Michael Sorell, and Feng Zhu. “Scale .1
Without Mass: Business Process Replication and Industry Dynamics.” HBS
working paper, September 2006 (http://www.hbs.edu/research/pdf/07-016.pdf;
.(also http://hbswk.hbs.edu/item/5532.html

Brynjolfsson, Erik, Lorin Hitt, and Heekyung Kim. “Strength in Numbers: .1
How Does Data-Driven Decisionmaking Affect Firm Performance?” ICIS
2011 Proceedings, Paper 13
(http://aisel.aisnet.org/icis2011/proceedings/economicvalueIS/13; also
.(available at http://papers.ssrn.com/so13/papers.cfm?abstracLid=1819486

.Byrne, John. The Whiz Kids. Doubleday, 1993 .2

Cate, Fred H. “The Failure of Fair Information Practice Principles.” In Jane .2
K.Winn, ed., Consumer Protection in the Age of the “Information Economy”
.(Ashgate, 2006), p. 341 et seq

Chin, A., and A. Klinefelter. “Differential Privacy as a Response to the .2
Reidentification Threat: The Facebook Advertiser Case Study.” 90 North
.(Carolina Law Review 1417 (2012

Crosby, Alfred. The Measure of Reality: Quantification and Western Society, .2
.1250-1600. Cambridge University Press, 1997

Cukier, Kenneth. “Data, Data Everywhere.” The Economist Special Report, .2
.February 27, 2010, pp. 1-14

.25

Tracking“._____
Social Media: The Mood of the Market.” Economist.com, June 28, 2012
(http://www.economist.com/blogs/graphicdetail/2012/06/tracking-social-
.(media

Davenport, Thomas H., Paul Barth, and Randy Bean. “How ‘Big Data’ Is .2
Different.” Sloan Review, July 30, 2012 (http://sloanreview.mit.edu/the-
.(/magazine/2012-fall/54104/how-big-data-is-different

Di Quinzio, Melanie, and Anne McCarthy. “Rabies Risk Among Travellers.” .2
.CMAJ 178, no. 5 (2008), p. 567

Drogin, Marc. Anathema! Medieval Scribes and the History of Book Curses. .2
.Allanheld and Schram, 1983

Dugas, A. F., et al. “Google Flu Trends: Correlation with Emergency .2
Department Influenza Rates and Crowding Metrics.” CID Advanced Access,
.January 8, 2012. DOl 1O.1093/cid/cir883

Duggan, Mark, and Steven D. Levitt. “Winning Isn’t Everything: Corruption .3
in Sumo Wrestling.” American Economic Review 92 (2002), pp. 1594-1605
.((http://pricetheory.uchicago.edu/levittjPapers/DugganLevitt2002.pdf

Duhigg, Charles. The Power of Habit: Why We Do What We Do in Life and .3
.Business. Random House, 2012

Duhigg, Charles. “How Companies Learn Your Secrets.” New York Times, .3
February 16, 2012
.((http://www.nytimes.com/2012/02/19/magazine/shoppinghabits.html

Dwork, Cynthia. “A Firm Foundation for Private Data Analysis.” .3
Communications of the ACM, January 2011, pp. 86-95
.((http://dl.acm.org/citation.cfm?id=1866739.1866758

Economist, The. “Rolls-Royce: Britain’s Lonely High-Flier.” The .3
.(Economist, January 8, 2009 (http://www.economist.com/node/12887368

.35
Building“._____
with Big Data: The Data Revolution Is Changing the Landscape of Business.”
The Economist, May 26, 2011
.(/(http://www.economist.com/node/18741392

.36
Official“._____
Statistics: Don’t Lie to Me, Argentina.” The Economist, February 25, 2012
.((http://www.economist.com/node/21548242

.3
Counting“._____
Every Moment.” The Economist, March 3, 2012
.((http://www.economist.com/node/21548493

.38
Vehicle“._____
Data Recorders: Watching Your Driving.” The Economist, June 23, 2012
.((http://www.economist.com/node/21557309

Edwards, Douglas. I'm Feeling Lucky: The Confessions of Google .3
.Employee Number 59. Houghton Mifflin Harcourt, 2011

Ehrenberg, Rachel. "Predicting the Next Deadly Manhole Explosion." .4
Wired, July 7, 2010 (http://www.wired.com/wiredscience/201O/07/manhole-
.(explosions

Eisenstein, Elizabeth L. The Printing Revolution in Early Modern Europe. .4
.Cambridge University Press, 1993

Etzioni, Oren, C. A. Knoblock, R. Tuchinda, and A. Yates. "To Buy or Not to .4
Buy: Mining Airfare Data to Minimize Ticket Purchase Price." SIGKDD '03,
August 24-27, 2003 (http://knight.cis.temple.edu/-yates//papers/hamlet-
.(kdd03.pdf

Frei, Patrizia, et al. "Use of Mobile phones and Risk of Brain Tumours: .4
Update of Danish Cohort Study." BMJ 2011, 343
.((http://www.bmj.com/content/343/bmj.d6387

Furnas, Alexander. "Homeland Security's 'Pre-Crime' Screening will Never .4
Work." The Atlantic Online, April 17, 2012
(http://www.theatlantic.com/technology/archive/2012/04/homeland-securitys-
.(/pre-crime-screening-will-never-work/255971

.Garton Ash, Timothy. The File. Atlantic Books, 2008 .4

Geron, Tomio. "Twitter's Dick Costolo: Twitter Mobile Ad Revenue Beats .4
Desktop on Some Days." Forbes, June 6, 2012
(http://www.forbes.com/sites/tomiogeron/2012/06/06/twitters-dick-costolo-
.(/mobile-ad-revenue-beats-desktop-on-some-days

Ginsburg, Jeremy, et al. "Detecting Influenza Epidemics Using Search .4
Engine Query Data." Nature 457 (2009), pp. 1012-14 (http://www.nature.com/
.(nature/journaljv457/n7232/full/nature07634.html

Golder, Scott A., and Michael W. Macy. "Diurnal and Seasonal Mood Vary .4
with Work, Sleep, and Daylength Across Diverse Cultures." Science 333
.(September 30, 2011), pp. 1878-81

Golle, Philippe. "Revisiting the Uniqueness of Simple Demographics in the .4
US Population." Association for Computing Machinery Workshop on Privacy
.in Electronic Society 5 (2006), pp. 77-80

Goo, Sara Kehaulani. "Sen. Kennedy Flagged by No-Fly List." Washington .5
Post, August 20, 2004, p. A01 (http://www.washingtonpost.com/wp-
.(dyn/articles/A17073-2004Aug19.html

Haeberlen, A., et al. "Differential Privacy Under Fire." In SEC'l1: .5
Proceedings of the 20th USENIX conference on Security, p. 33

.((http://www.cis.upenn.edu/-ahae/papers/fuzz-sec20ll.pdf

.Halberstam, David. The Reckoning. William Morrow, 1986 .5

Haldane, J. B. S. “On Being the Right Size.” Harper’s Magazine, March .5
.(/1926 (http://harpers.org/archive/1926/03/on-being-the-right-size

Halevy, Alon, Peter Norvig, and Fernando Pereira. “The Unreasonable .5
.Effectiveness of Data.” IEEE Intelligent Systems, March/April 2009, pp. 8-12

Harcourt, Bernard E. Against Prediction: Profiling, Policing, and Punishing .5
.in an Actuarial Age. University of Chicago Press, 2006

Hardy, Quentin. “Bizarre Insights from Big Data.” NYTimes.com, March 28, .5
2012 (http://bits.blogs.nytimes.com/2012/03/28/bizarre-insights-from-big-
.(/data

Hays, Constance L. “What Wal-Mart Knows About Customers’ Habits.” .5
New York Times, November 14, 2004
.((http://www.nytimes.com/2004/ll/14/business/yourmoney/14 wal.html

Hearn, Chester G. Tracks in the Sea: Matthew Fontaine Maury and the .5
.Mapping of the Oceans. International Marine/McGraw-Hill, 2002

Helland, Pat. “If You Have Too Much Data then “‘Good Enough’ Is Good .5
.Enough.” Communications of the ACM, June 2011, p. 40 et seq

Hilbert, Martin, and Priscilla Lopez. “The World’s Technological Capacity to .6
Store, Communicate, and Compute Information.” Science 1 (April 2011),
.pp.60-65

.61

How“.____
to Measure the World’s Technological Capacity to Communi- cate, Store and Compute Information?” International Journal of Com- munication (2012), pp. 1042-55 (ijoc.org/ojs/index.php/ijoc/article/
.(viewFile/1562/742

Holson, Laura M. “Putting a Bolder Face on Google.” New York Times, .6
March 1, 2009, p. BU 1
.((http://www.nytimes.com/2009/03/01/business/0lmarissa.html

Hopkins, Brian, and Boris Evelson. “Expand Your Digital Horizon with Big .6
.Data.” Forrester, September 30, 2011

Hotz, Robert Lee. “The Really Smart Phone.” Wall Street Journal, April 22, .6
2011
nline.wsj.com/article/SB10001424052748704547604576263261679848814.html

Hutchins, John. “The First Public Demonstration of Machine Translation: .6
The Georgetown-IBM System, 7th January 1954.” November 2005

.((http://www.hutchinsweb.me.uk/GU-IBM-2005.pdf

Inglehart, R., and H. D. Klingemann. Genes, Culture and Happiness. MIT .6
.Press, 2000

.Isaacson, Walter. Steve Jobs. Simon and Schuster, 2011 .6

Kahneman, Daniel. Thinking, Fast and Slow. Farrar, Straus and Giroux, .6
.2011

Kaplan, Robert S., and David P. Norton. Strategy Maps: Converting .6
Intangible Assets into Tangible Outcomes. Harvard Business Review Press,
.2004

Karnitschnig, Matthew, and Mylene Mangalindan. “AOL Fires Technology .7
.Chief After Web-Search Data Scandal.” Wall Street Journal, August 21, 2006

Keefe, Patrick Radden. “Can Network Theory Thwart Terrorists?” New York .7
Times, March 12, 2006
.((http://www.nytimes.com/2006/03/12/magazine/312wwlln_essay.html

Kinnard, Douglas. The War Managers. University Press of New England, .7
.1977

Kirwan, Peter. “This Car Drives Itself.” Wired UK, January 2012 .7
(http://www.wired.co.uk/magazine/archive/2012/0l/features/this-car-drives-
.(itself

Kliff, Sarah. “A Database That Could Revolutionize Health Care.” .7
.Washington Post, May 21, 2012

Kruskal, William, and Frederick Mosteller. “Representative Sampling, IV: .7
The History of the Concept in Statistics, 1895-1939.” International Statistical
.Review 48 (1980), pp. 169-195

Laney, Doug. “To Facebook You’re Worth $80.95.” Wall Street Journal, May .7
3, 2012 (http://blogs.wsj.com/cio/2012/05/03/to-facebook-youre-worth-80-
.(/95

Latour, Bruno. The Pasteurization of France. Harvard University Press, .7
.1993

Levitt, Steven D., and Stephen J. Dubner. Freakonomics: A Rogue .7
.Economist Explores the Hidden Side of Everything. William Morrow, 2009

.Levy, Steven. In the Plex. Simon and Schuster, 2011 .7

Lewis, Charles Lee. Matthew Fontaine Maury: The Pathfinder of the Seas. .8
.U.S. Naval Institute, 1927

Lohr, Steve. “Can Apple Find More Hits Without Its Tastemaker?” New .8
York Times, January 18, 2011, p. Bl
.((http://www.nytimes.com/20ll/01/19/technology/companies/19innovate.html

Lowrey, Annie. “Economists’ Programs Are Beating U.S. at Tracking .8
Inflation.” Washington Post, December 25, 20lO
(http://www.washingtonpost.com/wp-
.(dyn/content/article/2010/12/25/AR2010122502600.html
Macrakis, Kristie. Seduced by Secrets: Inside the Stasi’s Spy-Tech World. .8
.Cambridge University Press, 2008
Manyika, James, et al. “Big Data: The Next Frontier for Innovation, .8
Competition, and Productivity.” McKinsey Global Institute, May 2011 (http://
www.mckinsey.com/insights/mgifresearch/technology_and_innovation/
.(big_data_the_nexLfrontieLfoLinnovation
Marcus, James. Amazonia: Five Years at the Epicenter of the Dot.Com .8
.Juggernaut. The New Press, 2004
Margolis, Joel M. “When Smart Grids Grow Smart Enough to Solve .8
Crimes.” Neustar, March 18, 20l0
(http://energy.gov/sites/prod/files/gcprod/documents/
.(Neustar_Comments_DataExhibitA.pdf
Maury, Matthew Fontaine. The Physical Geography of the Sea. Harper, .8
.1855
Mayer-Schonberger, Viktor. “Beyond Privacy, Beyond Rights: Towards a .8
‘Systems’ Theory of Information Governance.” 98 California Law Review
.(1853 (2010
.—— .8
Delete: The Virtue of Forgetting in the Digital Age. Princeton University Press,
2nd ed.,
.2011
McGregor, Carolyn, Christina Catley, Andrew James, and James Padbury. .9
“Next Generation Neonatal Health Informatics with Artemis.” In European
Fed eration for Medical Informatics, User Centred Networked Health Care, ed.
.A. Moen et al. (lOS Press, 2011), p. 117 et seq
McNamara, Robert S., with Brian VanDeMark. In Retrospect: The Tragedy .9
.and Lessons of Vietnam. Random House, 1995
Mehta, Abhishek. “Big Data: Powering the Next Industrial Revolution.” .9
.Tableau Software White Paper, 2011
Michel, Jean-Baptiste, et al. “Quantitative Analysis of Culture Using .9
Millions of Digitized Books.” Science 331 (January 14, 2011), pp. 176-182
.((http://www.sciencemag.org/content/331/6014/176.abstract
Miller, Claire Cain. “U.S. Clears Google Acquisition of Travel Software.” .9

New York Times, April 8, 2011 (http://www.nytimes.com/20ll/04/09/
.(technology/09google.html?_r=0

Mills, Howard. “Analytics: Turning Data into Dollars.” Forward Focus, .9
December 2011 (http://www.deloitte.com/assets/Dcom-
es/Local%20Assets/Documents/FSI/US_FSLForward%20Focus_Analytics_Turn
.(ing’’1020data%20into%20dollars_1207ll.pdf

Mindell, David A. Digital Apollo: Human and Machine in Spaceflight. MIT .9
.Press, 2008

Minkel, J. R. “The U.S. Census Bureau Gave Up Names of Japanese- .9
Americans in WW II.” Scientific American, March 30, 2007
(http://www.scientificamerican.comjarticle.cfm?id=confirmed-the-us-census-
.(b

Murray, Alexander. Reason and Society in the Middle Ages. Oxford .9
.University Press, 1978

Nalimov, E. V., G. McC. Haworth, and E. A. Heinz. “Space-Efficient .9
Indexing of Chess Endgame Tables.” ICGA Journal 23, no. 3 (2000), pp. 148-
.162

Narayanan, Arvind, and Vitaly Shmatikov. “How to Break the Anonymity .10
of the Netfiix Prize Dataset.” October 18, 2006, arXiv:cs/0610105
.((http://arxiv.org jabsjcsj0610l05

101

Robust“.____

De-Anonymization of Large Sparse Datasets.” Proceedings of the 2008 IEEE
Symposium on Security and Privacy, p. 111(http://www.cs.utexas.edu/-
.(shmat/shmat_oak08netflix.pdf

Nazareth, Rita, and Julia Leite. “Stock Trading in U.S. Falls to Lowest .10
Level Since 2008.” Bloomberg, August 13, 2012
(http://www.bloomberg.com/news/20l2-08-13/stock-trading-in-u-s-hits-Iowest-
.(Ievel-since-2008-as-vix-falls.html

.Negroponte, Nicholas. Being Digital. Alfred Knopf, 1995 .10

Neyman, Jerzy. “On the Two Different Aspects of the Representative .10
Method: The Method of Stratified Sampling and the Method of Purposive
Selection.” Journal of the Royal Statistical Society 97, no. 4 (1934), pp. 558-
.625

Ohm, Paul. “Broken Promises of Privacy: Responding to the Surprising .10
.(Failure of Anonymization.” 57 UCLA Law Review 1701(2010

Onnela, J. P., et al. “Structure and Tie Strengths in Mobile Communication .10

Networks.” Proceedings of the National Academy of Sciences of the United States of America (PNAS) 104 (May 2007), pp. 7332-36 (http://nd.edu/- .(dddas/Papers/PNAS0610245104vl.pdf

Palfrey, John, and Urs Gasser. Interop: The Promise and Perils of Highly .10 .Interconnected Systems. Basic Books, 2012

Pearl, Judea. Causality: Models, Reasoning and Inference, 2nd ed. .10 .Cambridge University Press, 2009

President’s Council of Advisors on Science and Technology. “Report to the .10 President and Congress, Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology.” Decem ber 2010 (http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd- .(report-2010.pdf

Priest, Dana and William Arkin. “A Hidden World, Growing Beyond Con .11 trol.” Washington Post, July 19, 2010 (http://projects.washingtonpost.com/top- ,(/secret-america/articles/a-hidden-world-growing-beyond-control/print

Query, Tim. “Grade Inflation and the Good-Student Discount.” .11 Contingencies Magazine, American Academy of Actuaries, May-June 2007 .((http://www.contingencies.org/mayjun07/tradecraft.pdf

Quinn, Elias Leake. “Smart Metering and Privacy: Existing Law and .11 Competing Policies; A Report for the Colorado Public Utility Commission.” Spring 2009 .((http://www.w4ar.com/Danger_of_Smar_Meters_Colorado_Report.pdf

Reshef, David, et al. “Detecting Novel Associations in Large Data Sets.” .11 .Science (2011), pp. 1518-24

Rosenthal, Jonathan. “Banking Special Report.” The Economist, May 19, .11 .2012, pp.7-8

Rosenzweig, phil. “Robert S. McNamara and the Evolution of Modern .11 Management.” Harvard Business Review, December 2010, pp. 87-93 (http://hbr.org/2010/12/robert-s-mcnamara-and-the-evolution-of-modern- .(management/ar/pr

Rudin, Cynthia, et al. “21st-Century Data Miners Meet 19th-Century .11 .Electrical Cables.” Computer, June 2011, pp. 103-105

Machine“.____ Learning for the New York City Power Grid.” IEEE Transactions on Pattern Analysis and Machine Intelligence 34.2 (2012), pp. 328-345

.((http://hdl.handle.net/1721.1/68634

Rys, Michael. “Scalable SQL.” Communications of the ACM, June 2011, .11
.48, pp.48-53

Salathe, Marcel, and Shashank Khandelwal. “Assessing Vaccination .11
Sentiments with Online Social Media: Implications for Infectious Disease
Dynamics and Control.” PIOS Computational Biology 7, no. 10 (October
.(2011

Savage, Mike, and Roger Burrows. “The Coming Crisis of Empirical .12
.Sociology.” Sociology 41 (2007), pp. 885-899

Schlie, Erik, Jorg Rheinboldt, and Niko Waesche. Simply Seven: Seven .12
.Ways to Create a Sustainable Internet Business. Palgrave Macmillan, 2011

Scanlon, Jessie. “Luis von Ahn: The Pioneer of ‘Human Computation.’” .12
Businessweek, November 3, 2008
(http://www.businessweek.com/stories/2008-11-0 3/lu is-von-ahn-the-pioneer-
of-human-computa tion-businessweek-business-news-stock-market-and-
.(financial-advice

Scism, Leslie, and Mark Maremont. “Inside Deloitte’s Life-Insurance .12
Assessment Technology.” Wall Street Journal, November 19, 2010
nline.wsj.com/article/SB10001424052748704104104575622531084755588.html
124

Insurers“.____
Test Data Profiles to Identify Risky Clients.” Wall Street Journal, November
19, 2010 (http://online.wsj.com/article/SB100014240527487046
.(48604575620750998072986.html

Scott, James. Seeing Like a State: How Certain Schemes to Improve the .12
.Human Condition Have Failed. Yale University Press, 1998

Seltzer, William, and Margo Anderson. “The Dark Side of Numbers: The .12
Role of Population Data Systems in Human Rights Abuses.” Social Research
.68 (2001) pp. 481-513

Silver, Nate. The Signal and the Noise: Why So Many Predictions Fail- But .12
.Some Don’t. Penguin, 2012

Singel, Ryan. “Netflix Spilled Your Brokeback Mountain Secret, Lawsuit .12
Claims.” Wired, December 17, 2009
.(/(http://www.wired.com/threatlevel/2009/l2/netflix-privacy-lawsuit

Smith, Adam. The Wealth of Nations (1776). Reprinted Bantam Classics, .12
2003. A free electronic version is available
.((http://www2.hn.psu.edu/faculty/jmanis/adam-smith/Wealth-Nations.pdf

Solove, Daniel J. The Digital Person: Technology and Privacy in the .13
.Information Age. NYU Press, 2004

Surowiecki, James. “ABillion Prices Now.” New Yorker, May 30, 2011 .13
(http://www.newyorker.com/talk/financial/20ll/05/30/ll0530ta_talk_
.(surowiecki

Taleb, Nassim Nicholas. Fooled by Randomness: The Hidden Role of .13
.Chance in Life and in the Markets. Random House, 2008

.—— .13
The Black Swan: The Impact of the Highly Improbable. 2nd ed., Random House,
.2010

Thompson, Clive. “For Certain Tasks, the Cortex Still Beats the CPU.” .13
Wired, June 25, 2007 (http://www.wired.com/techbiz/it/magazine/15-
.(07/ff_humancomp?currentPage=all

Thurm, Scott. “Next Frontier in Credit Scores: Predicting Personal .13
Behavior.” Wall Street Journal, October 27, 2011
online.wsj.com/article/SB1000l424052970203687504576655l82086300912.html

Tsotsis, Alexia. “Twitter Is at 250 Million Tweets per Day, iOS 5 .13
Integration Made Signups Increase 3x.” TechCrunch, October 17, 2011
.(/(http://techcrunch.com/20ll/l0/l7/twitter-is-at-250-million-tweets-per-day

Valery, Nick. “Tech.View: Cars and Software Bugs.” The Economist, May .13
16, 2010
://www.economist.com/blogs/babbage/20l0/05/techview_cars_and.software.bugs

Vlahos, James. “The Department Of Pre-Crime.” Scientific American 306 .13
.(January 2012), pp. 62-67

Von Baeyer, Hans Christian. Information: The New Language of Science. .13
.Harvard University Press, 2005

von Ahn, Luis, et al. “reCAPTCHA: Human-Based Character Recognition .14
via Web Security Measures.” Science 321 (September 12, 2008), pp. 1465-68
.((http://www.sciencemag.org/content/321/5895/l465.abstract

Watts, Duncan. Everything Is Obvious Once You Know the Answer: How .14
.Common Sense Fails Us. Atlantic, 2011

Weinberger, David. Everything Is Miscellaneous: The Power of the New .14
.Digital Disorder. Times, 2007

Weinberger, Sharon. “Intent to Deceive.” Nature 465 (May 2010), pp. 412- .14
”.(415 (http://www.nature.com/news/20l0/l00526/full/4654l2a.html

Terrorist“.______
‘Pre-crime’ Detector Field Tested in United States.” Nature, May 27, 2011
.((http://www.nature.com/news/20ll/110527/full/news.20l1.323.html
Whitehouse, David. “UK Science shows Cave Art Developed Early.” BBC .14
News Online, October 3, 2001
.((http;//news.bbc.co.uk/l/hi/sci/tech/1577421.stm
Wigner, Eugene. “The Unreasonable Effectiveness of Mathematics in the .14
Natural Sciences.” Communications on Pure and Applied Mathematics l3, no.
.1 (1960), pp.I-14
.Wilks, Yorick. Machine Translation: Its Scope and Limits. Springer, 2008 .14
Wingfield, Nick. “Virtual Products, Real Profits; Players Spend on Zynga’s .14
Games, but Quality Turns Some off.” Wall Street Journal, September 9, 2011
online.wsj.com/article/SBI0001424053111904823804576502442835413446.html

شكر وتقدير

لقد كنا محظوظين للعمل مع واحد من أوائل عمالقة مجال شبكات وابتكارات المعلومات لويس إم. برانسكومب والتعلم منه، والذي ظلت فطنته وفصاحته واحترافيته وذكاؤه وشغفه وطاقته التي لا حدود لها تلهمنا حتى الآن. وكذلك شريكته الودودة والحكيمة كوني مولين، والتي نأسف لها على عدم انصياعنا لاقتراحها بأن نجعل عنوان الكتاب "البيانات الخارقة".

كان مؤمن مالك، مساعد باحث ممتازًا يتميز بالفطنة والاجتهاد. لقد تشرفنا بأن يقدمنا كل من ليزا آدامز ودافيد ميللر الرائعين من وكالة "جاراموند"، في كل ما يعملان. كان إدوارد مولان، محرر هذا الكتاب، رائعًا ـــ ممثلاً عن طائفة نادرة من المحررين الذين يعرفون جيدًا كيف يحررون النصوص ويتحدون تفكيرنا، لذا كانت النتيجة أفضل مما أملنا بمراحل. إننا نقدم الشكر إلى جميع العاملين في هوتن ميفلين هاركورت، خاصة بيث بيرلييه فولر وبن هيمان. وكذلك كاميل سميث على التدقيق اللغوي الاحترافي. نشكر أيضًا جيمس فرانشام من جريدة **ذا إكونوميست** على تقصيه الممتاز للحقائق ونقده البارع لمسودة الكتاب.

نوجه الشكر بشكل خاص لجميع العاملين في مجال البيانات الضخمة والذين قضوا معنا بعضًا من وقتهم في شرح عملهم، وخصوصًا أورين إتزيوني، وسينثيل رودين، وكارولين ماكجريجور ومايك فلاورز.

بالنسبة للشكر والتقدير من فيكتور: أقدم بالغ شكري إلى فيليب إيفانز الذي طالما كان فكره متقدمًا على الجميع بخطوتين وعبّر عن أفكاره بدقة وبلاغة، وعن مناقشاته التي استمرت معي فترة تجاوزت عقدًا من الزمان.

أقدم شكري أيضًا لشريكي السابق دافيد لازر، الذي كان أحد أساتذة البيانات الضخمة الأوائل الأقوياء، والذي أخذت بنصائحه عدة مرات من قبل.

أقدم شكري إلى المشاركين في حوار البيانات الرقمية بجامعة أوكسفورد عام 2011 (الذي كان يتمحور حول البيانات الضخمة)، وخاصة رئيسه المشارك فريد كايت على مناقشاته القيمة.

لقد قدم معهد أوكسفورد للإنترنت، الذي أعمل به، المناخ المناسب لتأليف هذا الكتاب، حيث إن الكثير من زملائي انخرطوا في إجراء أبحاث عن البيانات الضخمة. لم أتمكن من التفكير في مكان أفضل منه لأقوم بتأليف هذا الكتاب. أقدر أيضًا بامتنان دعم كلية كيبل، التي أتشرف بزمالتها. بدون هذا الدعم الذي حصلت عليه، لم أكن لأتمكن من الوصول إلى بعض المصادر المهمة الرئيسية التي استخدمتها في هذا الكتاب.

عادة ما تدفع الأسرة الثمن الأكبر عندما يقوم المرء بتأليف كتاب ما، حيث إنني لم أكن أقضي ساعات طويلة أمام شاشة الحاسب فحسب، أو في مكتبي، بل أيضًا بسبب الساعات الطويلة التي كنت موجودًا فيها في المنزل بجسدي فقط ولكن كان عقلي مشغولاً بالأفكار، لذا فإنني أطلب العفو والغفران من زوجتي برجيت ومن طفلي فيكتور الصغير، وأعدهم بأنني سوف أحاول بجد أكبر.

بالنسبة لشكر وتقدير كين الشخصيين: أشعر بالامتنان للكثير من علماء البيانات العظام الذين ساعدوني، خاصة جيف هامرباخر، وعمرو عوض الله، ودي جي باتل، ومايكل دريسكول، ومايكل فريد والكثير من الزملاء من شركة جوجل الذين ساعدوني على مدار سنوات طوال (من بينهم هال فاريان، وجيريمي جينسبرج، وبيتر نورفيج، وأودي مانبر، وغيرهم، وفي الوقت ذاته كانت المحادثات القصيرة مع كل من إريك شميدت ولاري بايج لا تقدر بثمن).

لقد أثري تفكيري بواسطة تيم أورايلي عبقري عصر الإنترنت. وأيضًا مارك بينيوف من شركة Salesforce.com، والذي يعتبر معلمي. وكانت أفكار ماتيو هيندمان لا مثيل لها، كالعادة. وقدم لي جايمس جوسزكزا من ديلويت، مساعدة رائعة، كما فعل جيوف هيات، الصديق القديم ورائد البيانات المتسلسلة. وأعبر عن شكري الخاص إلى بيت واردن، الذي يعتبر فيلسوفًا وممارسًا للبيانات الضخمة.

قدم لي الكثير من الأصدقاء الأفكار والنصائح، من بينهم جون ترنر، وأنجيليكا وولف، ونيكو وايسش، ودافيد ويشارت، وأنا باتيريك، وبلاين هاردن وجيسيكا كوال. وقد ألهمني آخرون بالموضوعات الرئيسية في هذا الكتاب من بينهم بلايز أجويرا وأركاس، وإريك هورفيتز، ودافيد أويرباخ، وجيل إلباز، وتايلر بيل، وأندرو ويكوف والكثيرون من العاملين في منظمة التعاون والتنمية الاقتصادية، وستيفن بروبست وفريق العمل بشركة تيراداتا، وأنتوني جولدبلووم وجيريمي هوارد من شركة كاجل، وإد دامبل، وروجر ماجولاس وفريق العمل بشركة أوريلي ميديا، وإدوارد لازوسكا. إن جايمس كورتادا أحد عظماء هذا المجال. أقدم الشكر أيضًا لكل من بينج لي من شركة إكسل بارتنرز وروجر إهرنبرج من شركة آي إيه فنشرز.

قدم لي زملائي في جريدة ذا إيكونوميست أفكارًا رائعة ودعمًا لا نهائيًّا. أقدم شكرًا خاصًّا للمحررين توم ستانداج، ودانيال فرانكلين، وجون ميكلتوايت، بالإضافة إلى باربرا بك والتي حررت التقرير الخاص تحت عنوان "البيانات، البيانات في كل مكان"، الذي يعتبر أصل هذا الكتاب. وأشكر هنري تريكس ودومينيك زيجلر، زميليّ من طوكيو، واللذين كانا بمثابة القدوة التي تمثلت بها فيما يتعلق بالبحث عن جديد وعرضه بشكل جميل. قدم أوليفر مورتون نصائحه المعتادة عندما كنت في أمسّ الحاجة إليها.

لقد قدم لي مؤتمر سالزبرج الدولي في النمسا، التركيبة المثالية للفكر الشاعري والبحث الفكري الذي ساعدني على الكتابة والتفكير. وأمدني مؤتمر معهد أسبن في يوليو عام 2011 بالكثير من الأفكار، ولهذا أقدم الشكر للمشاركين والمنظم، تشارلي فايرستون. وكذلك أقدم بالغ شكري وتقديري إلى تيري إلينسكي على دعمها الرائع.

قدمت لي فرانسيس كايرنكروس، عميدة كلية إكستير بأكسفورد، مكانًا هادئًا للإقامة وتشجيعًا منقطع النظير، والتي لن يمكنني أن أوفيها حقها فيما يتعلق بالأسئلة عن التكنولوجيا والمجتمع الذي بني على أساسها، والتي طرحتها منذ خمس عشرة سنة مضت في كتابها The Death of Distance، العمل الذي ألهمني كصحفي مبتدئ. كان من المرضي لي أن أمر كل يوم في فناء

كلية إكستير عالمًا بأنني قد آمر بشعلة قامت هي بحملها من قبل، رغم أن الشعلة كانت أكثر سطوعًا في يدها.

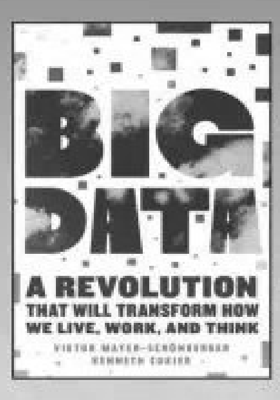
أقدم أعمق تقديري إلى أسرتي، التي احتملتني ـــ أو احتملت غيابي. يستحق والداي وأختي وغيرهم من أقاربي جزيل الشكر، ولكنني أحتفظ بأعمق امتناني إلى زوجتي هيذر وطفلينا تشارلوت وكاز، الذين دون دعمهم وتشجيعهم وأفكارهم، لم يكن ممكنًا أن أنهي هذا الكتاب.

نقدم كلانا جزيل الشكر إلى كثير من الأشخاص الذين ناقشوا معنا الموضوع الرئيسي للبيانات الضخمة، قبل حتى أن يصبح هذا المصطلح شائعًا. عند هذه النقطة، نقدم جزيل الشكر إلى المشاركين الدائمين على مدار أعوام في مؤتمر ريوشليكون عن سياسة المعلومات والذي كان فيكتور أحد المشاركين في تنظيمه، وكان كين مقرر لجنته. نقدم شكرًا خاصًّا إلى جوزيف الهادف، وبرنارد بنهامو، وجون سيلي براون، وهربرت بوركتر (الذي عرفنا على القبطان موراي)، وبيتر كولن، وإد فلتين، وإرس جاسر، وجوي إتو، وجيف جوناس، ونيكلاس لاندبلاد، ودوجلاس ميريل، وريك موراي، وجوري أوندرجكا، وبول شوارتز.

فيكتور ماير شونبرجر

كينيث كوكيير

أوكسفورد/ لندن، أغسطس 2012



"في كل قرن، تظهر مجموعة من الكتب التي تغير من أسلوب تفكيرنا في كل شيء. ويعتبر هذا الكتاب أحد هذه الكتب. وقد أصبح المجتمع يتطلع إلى التغيير الذي سيحدثه هذا الكتاب. إنه فعلًا بداية عظيمة للتغيير".

-لورانس ليسيج، أستاذ كرسي "روي إل. فورمان" للقانون بكلية الحقوق بجامعة هارفارد، ومؤلف كتابي *Remix* و *Free Culture*

"يطرق هذا الكتاب مجالاً جديدًا لتحديد الكيفية التي يحول بها الكم الهائل من المعلومات طريقة فهمنا الأساسية للعالم من حولنا. هذا الكتاب، الذي يطرح جدلاً جريئًا وأسلوبًا رائعًا، يظهر بوضوح كيف يمكن للشركات أن تحرر قيمها، وكيف يحتاج صناع القرار لأن يكونوا حذرين، وكيف أن الأنماط الإدراكية للجميع تحتاج إلى التغيير".

- جوي إتو، مدير معمل إم آي تي للإعلام

"يجب أن يقرأ هذا الكتاب كل من يرغب في أن يبقى على رأس أحد أهم التوجهات التي ستحدد مستقبل عالم الأعمال".

- مارك بينيوف، رئيس مجلس الإدارة والمدير التنفيذي لموقع salesforce.com

"نظرة متفائلة وعملية على ثورة البيانات الكبرى - الكتاب الذي سيجعلك تفكر بعمق في التغيرات العظيمة التي تقوم بها والتغيرات الأعظم التي ستقوم بها"

- كوري دوكتورو، boingboing.com

"كما تتميز المياه في مجملها بخاصية البلل في حين لا يمتلك جزيء الماء بمفرده هذه الخاصية، فإن تكتل البيانات يمكنه أن يكشف عن كم من المعلومات لا يمكن لقصاصات فردية من البيانات كشفه. يكشف لنا المؤلفان الطرق المدهشة التي يمكن لكم كبير ومعقد من البيانات أن يستخدمها لتوقع كل شيء بدءًا بأنماط التسويق وحتى تفشي مرض الأنفلونزا"

- كلاي شيركي، مؤلف كتابي *Cognitive Surplus* و *Here Comes Everybody*

"يزيل هذا الكتاب الرائع الغموض والمبالغة اللذين يحيطان بالبيانات الكبرى، ويجب أن يقرأه كل من يعمل في مجالات الأعمال وتكنولوجيا المعلومات والسياسة العامة والاستخبارات والطب، وأي شخص آخر يشعر بالفضول بشأن ما سيحدث في المستقبل"

- جون سيلي براون، كبير علماء سابق بشركة زيروكس، ورئيس مركز أبحاث زيروكس في بالو ألتو

"يزخر هذا الكتاب برؤى ثاقبة عن الطرق الحديثة لترويض المعلومات ويعرض رؤية مقنعة للمستقبل. ويجب أن يقرأه كل من يستخدم أو يتأثر بالبيانات الكبرى"

- جيف جوناس، زميل شركة آي بي إم، وكبير علماء تحليل الكيانات بها


ISBN 628-1072-07-613-9
6 281072 076139
282205439



